Creating an agent-based artificial market

Per Berséus

Master's thesis
Lunds Tekniska Högskola
Supervisors: Thomas Guhr and Rudi Schäfer

April, 2007

Abstract

This thesis is a study within econophysics, a research field where financial problems are investigated using methods from statistical physics.

An artificial stock market was implemented in Java and used to assess some assertions made in empirical studies by J.-P. Bouchaud and collegues. The model was able to reproduce several characteristics of high-fequency data from financial markets, e.g. a leptocurtic return distribution and a nearly constant return function. Much of the emphasis of this report is made on the methodology itself, since many modelling issues remain to be solved in the agent-based computational economics field. However, the market was also used to simlulate trading scenarios, where the impact of different trading strategies was analysed.

More specifically, it was shown in simulation that an overall diffusive price development can result from a competition between two populations of traders: one of liquidity providers, who create persistence in the prices, and another of liquidity takers, who create anti-persistence. It was also shown that a slowly decaying trade sign autocorrelation function could result from similar trading criteria, but not in any obvious way from liquidity provider's division of large market orders into several smaller ones.

Keywords: *agent-based modelling, ABM, artificial markets, complex systems, diffusion, econophysics, liquidity takers, liquidity providers, order book*

Contents

1 Introduction

1.1 Econophysics

This thesis applies concepts and methods from statistical physics on a financial setting. The blending of economics and physics is a fast-growing field of research, often referred to as *econophysics*.

1.1.1 Statistics on financial time series

For many reasons, *financial time series* have received much attention in recent years. Trades data have been recorded at the world's financial markets for more than two centuries, and at today's electronic markets every single action can be recorded and stored. Although such financial time series may concern different assets, different regions and different epochs, they present a rich statistical structure that seems to be universal [8]. The scientific exploration of these records of social human interactions is a challenge where many fundamental questions still remain unresolved.

Ever since Bachelier [2] investigated the properties of prices at stock markets, it is known that price changes follow, to a first approximation, a *random walk* (or *Brownian motion*) behaviour. As an important consequence, you cannot predict what the future price of an asset will be through a statistical analysis of past prices. However, as Bachelier remarks himself, in lack of *deterministic predictability* you can still rely on *statistical predictability* in financial markets. When millions of euros are invested, it is of course very important to know what the chances are of making money, as well as the risks of losing money.

1.1.2 Complexity of financial data

The great potential of the Brownian motion approach was manifested in the work of Black and Scholes [4] and it has become standard in financial theory. However, a closer look at the empirical data reveals a much richer structure than a simple random walk. For example, the distribution of price changes has much fatter tails than Gaussian (a fat-tailed or *leptokurtic* distribution is at hand when occurrences far from the mean are more common than would be expected from a normal distribution; details are given in section 2.2). To faithfully describe empirical data, more sophisticated models are needed.

As a matter of fact, it may be necessary to look deeper into the mechanisms of financial markets in order to understand the full complexity of price changes. This thesis is using an *agent-based model*, which on a very concrete level tries to mimic what actually happens when traders meet on a market, where money and assets change hands, where people try to hide their agendas, where trading strategies are carried out.

1.2 Agent-based modelling

1.2.1 Agent-based modelling: a bottom-up approach

In the modelling of financial markets, two fundamentally different approaches are commonly used. The *top-down* models, which are using a set of differential equations to describe fluctuating stock prices, managed early on to reproduce some of the most important features of real financial data. In several aspects, though, the necessary assumptions for such models have been proved not to correspond to the real world, and a *bottom-up* approach, such as the **agent-based modelling** (ABM) used in this thesis, seems more promising in many respects. The idea is to simulate the trading situation in detail: traders place orders to buy or sell, and a price curve emerges as a result of these actions.

The decisions of traders at a *microscopic* level will thus determine the *macroscopic* behaviour of the system. Macroscopic variables, such as the price of an asset, are not modelled as such; they are only the result of the combined micro-level actions of many agents. The researcher controls the initial state of the market and chooses the rule sets of the agents, but lets the market evolve by itself once the simulation has started.

One of the benefits of agent-based modelling is the clarity and transparency of the models. The human mind cannot immediately relate a set of differential equations to a bunch of traders yelling at each other at the trading floor, and many scientific models indeed suffer from such uncanny lack of transparency. By contrast, an agent-based model mimics in detail what is actually happening in the real world, which has a nice reassuring feel to it.

1.2.2 The early days of agent-based modelling

The agent-based approach can be very consuming in terms of computational power, and it did not flourish until the speed of microcomputers evolved. The first applications were not within finance: agent-based modelling has a strong intuitive appeal when applied to any biological or social system that involves a lot of actors. An early example of agent-based modelling is John Conway's *Game of Life* (see [11] for a nice online description), created in 1970, while LeBaron [19] quotes six studies from the mid-90's as early landmarks of agent-based modelling in finance. Amongst these is the renowned *Santa Fe Artificial Stock Market*, the most ambitious of all artificial markets (for reviews, see e.g. [20] and [18]).

1.2.3 Agent-based modelling in economics

The branch of agent-based modelling that deals with economic environments is sometimes referred to as *agent-based computational economics* (ACE) or *agent-based computational finance*, which naturally includes agent-based artificial markets. Even to summarize the research on this fragment of agent-based modelling is a little beyond the scope of this report. As pointed out in [12], an agent-based approach to financial markets draws in itself on several distinct literatures: the market microstructure literature, the experimental markets literature and the simulated markets literature. For general reviews, see for example [25] and [21].

Many researchers prefer not to impose any trading strategies on the agents, but are instead using for example genetic algorithms in an evolutionary approach, where agents gradually learn to use more efficient strategies [25]. These artificial markets would sort under the *adaptive complex systems* field. As pointed out by Tesfatsion [24], using fixed decision rules and creating representative agents is actually a top-down construction, and agent-based markets where learning is involved are thus more faithful to the bottom-up philosophy. However, they cannot answer questions on what influene a specific trading strategy would have on the market development. Therefore, fixed decision rules are sometimes preferred, as in this thesis.

Some articles take on a practical approach and address some of the considerations and difficulties that often need attention from anyone trying to create a new agent-based artificial market (e.g. [18] and especially [1]). This thesis report is trying to contribute in a similar way, laying as much weight on the modeling work as on the actual results.

1.3 Concepts of order-book trading

1.3.1 Financial markets

A *financial market* is a mechanism that allows traders to trade money for *securities* or *commodities*. Different markets deal with different such *assets* (*shares*, *derivatives*, raw materials) but they all share many common features. For example, if you want to sell something, you will have to find a buyer. In many markets, the offers of buyers and sellers are handled in a centralised manner, so that they are more easily matched. A common solution is to let all traders state their offers as orders and send them to an *order book*.

1.3.2 Order-book trading

A trader that wants to make an offer to sell or to buy at a specific price sends a *limit order* to the order book, where it is visible for all the other traders. If a market participant wants to buy or sell right away, she places a *market order* instead. This order has no price limit, but is matched with a suitable offer in the order book so that a trade will take place.

The order book is sorted, so that the buy order offering the highest upper price limit, the *best buy*, will be the one that is matched with the first incoming market sell order. Equally, an incoming market buy order will be matched with the *best ask*, that is the lowest price that a seller is asking for. If a market order requests a greater quantity (or *volume*) of shares than is offered by the best-priced limit order, the trade will also encompass limit orders deeper into the order book, with perhaps another price.

If several limit orders offer the same price, the earliest one will be matched before the later ones. This sorting principle is called *price-time priority*.

1.3.3 Stock exchanges

At a *stock exchange*, traders gather to buy or sell shares. The trading at stock exchanges is often, as in most of today's markets, handled electronically. A trader who wants to buy a share must place an electronic buy order, either a market order, where the trader agrees to buy at current market price (i.e. the best price any seller is offering), or a limit order, where the trader specifies which price she is prepared to pay. All limit orders, both buy and sell orders, are placed in the order book, so that everyone can see what the best current offers are.

1.3.4 An example of financial data

Bouchaud et al. [7] use *high-frequency data*, with time stamps accurate to the second, from the Paris Bourse. For each stock, there is a list of all successive *quotes*, which shows the state of the order book at all times. A quote consists of the best bid, b, the best ask, a, the available volume and a time stamp. A new quote is generated whenever the order book changes, for example as a result of a trade, a new limit order or the cancellation of an order from the order book.

There is also a list of successive *trades*, where the traded price and the traded volume are specified along with a time stamp. The data can be arranged so that every trade is preceded by at least one quote. The *sign of a trade* is then easily defined: if the traded price is above the last midpoint $m = (a + b)/2$, this trade will be assigned a sign value of $\varepsilon = +1$. This happens when the trade is triggered by a market order to buy, since the limit order that determined the price must have been at the upper end of the bid-ask spread, where limit sell orders are found. If the traded price is below the last midpoint price, the trade was triggered by a market order to sell, and the trade will be assigned a sign value of $\varepsilon = -1$.

A detailed description of the order book trading at the French stock exchange, better known as the *Paris Bourse*, can be found in the empirical study [3].

1.3.5 Bid-ask spread and tick size

The *bid-ask spread*, s, is the difference between the best ask, a, and the best bid, b:

$$s(t) = a(t) - b(t)$$

It shows how much you lose if you buy a share at market price and then immediately sell it at market. For shares that are traded a lot, the spread is generally small, sometimes down to a single *tick*. The *tick size* of a market determines how small the discrete price steps can be, just as we are used to rounding off prices at the local mall to what can be paid with the available coins of our local currency.

1.3.6 Liquidity and different types of traders

Liquidity on a stock market means that shares are traded continuously, so that you can count on someone actually selling shares to you as soon as you are willing to pay a reasonable price. On a *liquid market*, the spread between the best bid price and the best ask price is narrow enough for traders to easily agree on a price for a specific trade. Therefore, the spread can be used as a coarse measure of liquidity of a certain asset.

Liquidity is important for the well-functioning of a market. What actually creates liquidity are the limit orders. An actor that undertakes to emit limit orders and to make sure that the bid-ask spread is narrow at all times is traditionally called a *market maker*. At today's electronic markets, all participants can place limit orders, thereby acting as **liquidity providers**; they sometimes receive a bonus from the market organizers for doing so. Market participants who place market orders are instead acting as **liquidity takers** (the terminology is a bit fuzzy around these different types of traders and there exist other definitions than the ones used in this thesis).

Liquidity takers are sometimes called *informed traders*. Traders are informed if they have access to information that puts them in a better position than other market participants for predicting future prices. The information can come from many different sources, such as a *fundamental analysis* of the firm, statistical analyses of trade patterns, observations of macro-economic trends, etcetera. Unfortunately, informed traders rarely interpret their information correctly, and success rates of for example systematic hedge funds are typically around 52% (according to [8]).

If liquidity takers are right and really do anticipate price changes correctly, they earn money at the expense of liquidity providers. They are, however, often wrong, in which case they lose the bid-ask spread to the liquidity providers on closing their position. Besides, liquidity takers are not always acting on perceived information advantage. Their trades can also be due to hedging (risk-handling), mistakes, etcetera.

1.3.7 Tactical interplay between liquidity takers and liquidity providers

Bouchaud et al. [7] describes the interplay between liquidity providers and liquidity takers as characterised by an interesting logic:

1. Liquidity providers have reason to be careful. If they sell large volumes to a truly informed trader right before a price raise, it will be a costly mistake. Therefore, liquidity providers will be tempted to increase the price of their offer if someone starts buying. They will not increase the price too much, though, since they will then loose money if the price fluctuates back down to its initial value before they close their position.

2. Liquidity takers know that liquidity providers are suspicious. When liquidity takers want to buy large holdings, they do not want to appear on the radar screens and trigger an increase in the ask price. Therefore, they divide large orders into several smaller ones and disperse these over time, a tactical maneuvre that is known to be common behaviour in real markets.

Bouchaud et al. claim that this mechanism, where liquidity takers split large orders into smaller parts, explains the observed *temporal correlations* in the *sign of the trade* ($\varepsilon = +1$ for a market buy order, $\varepsilon = -1$ for a market sell order; see section 5.2.4). The existence of correlations complicates the picture of the price curve as a random walk: if the direction of a step at time t is correlated to the direction of the step at time $t-1$ or other preceding instances, the walk cannot be completely random.

1.4 Trading dynamics

1.4.1 Diffusiveness as a balance between opposing forces

Bouchaud et al. [7] argue that the seemingly random-walk nature of stock prices may in fact result from a fine-tuned interplay between two opposing tendencies: the correlated market orders of liquidity takers that lead to *persistence* and the mean-reverting limit orders of the liquidity providers that lead to *anti-persistence*. They see analogies with other complex systems that are driven by opposing forces, such as the heart (where the sympathetic system competes with the para-sympathetic system) or the human task of balancing a long stick. A small deviation from the critical point may in these cases lead to strong instabilities, and the authors suggest that this could be the right place to search for explanations to the *fat tails* (high frequency of extreme events, see section 2.2) and *volatility clustering* (i.e. when high volatility tends to be followed by more high volatility, see section 2.1 for an explanation of the term "volatility") observed in financial data.

In their analysis, Bouchaud et al. sometimes choose to take on the physicists view by looking at the price fluctuations as a *diffusive process*. This does not change the statistical description of the problem, as diffusion is mathematically described by the Brownian motion. In diffusion, the variance increases linearly with time, and we know that the variance of price differences is indeed proportional to the time interval considered. If there is persistence in the price so that it fluctuates less than this, a physicist would say that it must be affected by a *sub-diffusive* force. Equally, if the price fluctuations are stronger, there must be a *super-diffusive* force at work, creating anti-persistence. If both kinds of forces are present, the interplay between them could result in a process that is diffusive on average, but where perturbations can lead to severe instabilities.

Bouchaud et al. suggest that the population of liquidity providers represent a sub-diffusive force. As explained in the previous section, these traders make their money on the small difference between the bid price and the ask price. Since price fluctuations could make them loose this gain, they would prefer a constant stock price. According to Bouchaud et al. the liquidity providers therefore act in a way that preserves the price.

Liquidity takers, on the other hand, do not want a constant price in the long run, but they do want to be able to buy or sell a large holding without price adjustments on the market. They believe that they have access to better information than the market in general, and that they can make a profit for example by buying shares at a low price and selling these shares later on at a higher price. However, buying too many shares at once draws attention and also makes other traders less willing to sell at a low price, if they suspect that the trader who wants to buy actually knows something.

There is empirical evidence that liquidity providers actually make sure that the available volume in the order book represents just a small fraction of the market capitalization (the total value of all the issued stock), so that informed traders cannot buy or sell large volumes before prices can be adjusted [30]. Thus, the liquidity takers are forced to divide their trades into several smaller orders, *thereby creating correlations* in the order book.

1.4.2 Markovian and non-Markovian agents

In a *Markov process*, the future states depend only on the current state of the process and not on any past states. In other words, the process is *conditionally independent* of the past states, given the current state [28].

The Brownian motion is an example of a Markov process, and since price changes at a stock exchange follow this motion, we would expect decisions affecting future prices to be independent of the past states of the market. However, as described in sections 1.3.7 and 1.4.1, traders (more precisely liquidity takers) do sometimes decide on making a series of trades, spread out in time. If they place an order at a certain time t, but actually made the trading decision already at, say, time $t-2$, a non-Markovian component will be introduced into the system. The correlations created by this non-Markovian behaviour must be compensated for by other factors if the Brownian motion should stay a Brownian motion. The claim of Bouchaud et al. [7] that the strategies of liquidity providers constitute this compensating force, is challenged by Lillo and Farmer [22], who are suggesting that liquidity fluctuations are more likely to provide this effect. In [9], Bouchaud et al. convincingly defend their original assertions with evidence from empirical data.

1.5 A quick guide to this report

This master's thesis report accounts for the development of an artificial market with a special focus on the actual modelling and implementation. The target group is people who take an interest in the more technical aspects of agent-based modelling. It may be master students that are thinking about doing a similar project. It may be researchers that already are in the complex adaptive systems field and who wish to compare another working agent-based model with their own.

- In the introduction, **section 1**, some basic concepts have been introduced along with a minor survey of some research fields of relevance.
- The theoretically oriented **section 2** only briefly outlines the most important statistical background, without proofs or thorough explanations.
- In **section 3**, the purpose and relevance of this thesis is treated. Although the essential part for understanding this report is perhaps covered in subsection 3.1.2, the whole of section 3 tries to answer on a larger scale the important questions of why the research is meaningful.
- It is assumed that most people reading this report take at least some interest in the technical aspects of the model. Therefore, a substantial part of the report concerns pure implementation and modelling issues, in **section 4**.
- **Section 5** concerns with data treatment.
- The results are presented and discussed in **sections 6 and 7**.
- Some important results are summarized in a short conclusion in **section 8**.
- Contact information can be found in **section 9**.
- **Section 10** is the acknowledgements section.
- For the reader's convenience, a glossary of financial, modelling and statistical terms can be found in **appendix A**.

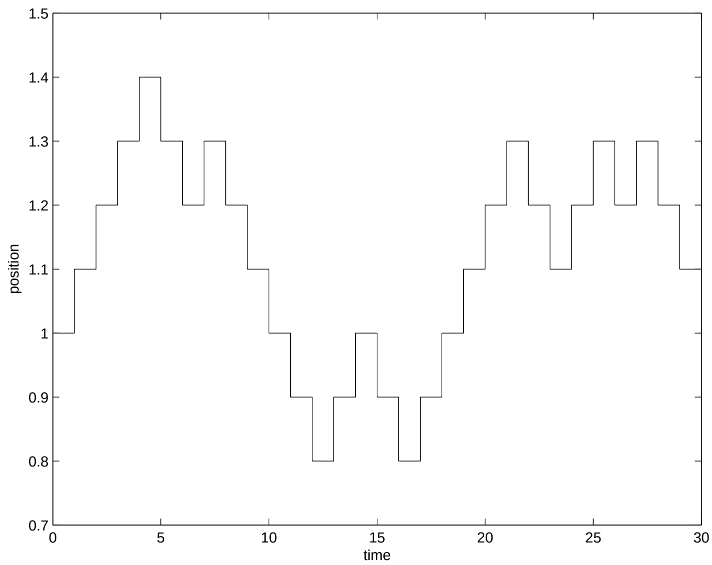

Figure 1: A random walk.

2 Theory

Much of the theroy in sections 2.1 and 2.1.1 is taken from [17], where proofs and a more elaborate description can be found.

2.1 Random walk and Brownian motion

A **random walk** is a stepwise movement, where each step is taken in a randomly chosen direction. An example of a random walk is an individual moving sideways on a straight line, taking steps either to the left or to the right, with equal probabilities. Although this individual's position cannot be predicted in a deterministic manner, it is possible to make statistical assertions. For example, the best prediction of the individual's future position would be exactly where she started from. A random walk is depicted in figure 1.

Brownian motion is the continuous equivalent to a random walk, see figure 2. If the step size of a random walk is diminished and tends towards zero, the walk approaches a Brownian motion. Einstein brought the attention of physicists to the Brownian motion when he treated it mathematically in 1905 [14], and several phenomena in nature are characterized by this type of motion: one well-known example is the movements of a particle immersed in a fluid. The Brownian motion has always been of interest when looking at financial data, since the price curve emerging from a financial market actually follows a random walk, at least to the first approximation.

Mathematically, the differential

$$dS = \sigma \eta \sqrt{dt} \tag{1}$$

describes the Brownian motion, where η is a dimensionless random number. The presence of the square root of the time differential makes this *stochastic*

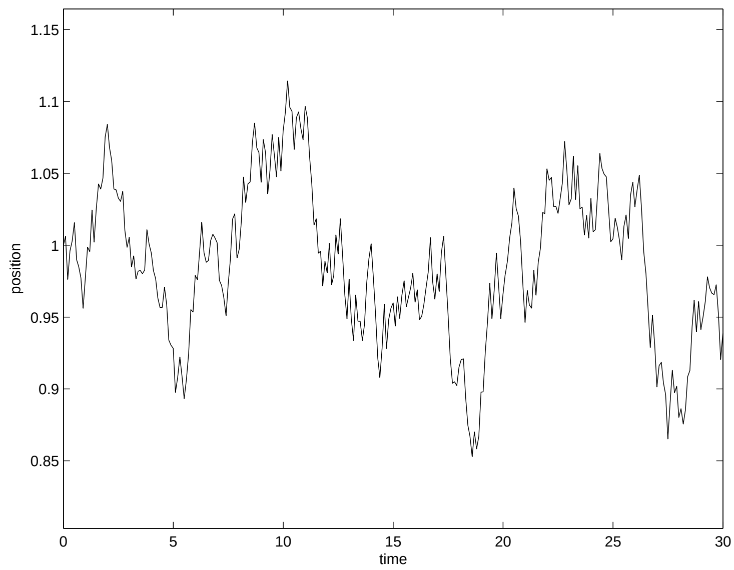

Figure 2: A Brownian motion.

differential equation quite different from the standard differential equations, which are sometimes referred to as *deterministic*. In statistics, the term *Wiener process* is preferred, since the Brownian motion is a type of stochastic process.

A *drift term* can be added

$$dS = \mu dt + \sigma\eta\sqrt{dt} \tag{2}$$

which means that the expected position will change linearly with time.

The Brownian motion, with or without drift, is a special case of the *Itō process*, described by

$$dS = A(S,t)\mu dt + B(S,t)\sigma\eta\sqrt{dt} \tag{3}$$

where A and B are arbitrary smooth coefficient functions of S and t.

Choosing $A = B = S$ yields the *geometric Brownian motion with a drift*:

$$dS = S\mu dt + S\sigma\eta\sqrt{dt} \tag{4}$$

which plays an important role in financial applications. Written in the form

$$\frac{dS}{S} = \mu dt + \sigma\eta\sqrt{dt} \tag{5}$$

both sides of the equation are dimensionless, and the left side is interpreted by economists as the *return*, that is how much you have earned compared to the investment you have made. The return can be expressed as Brownian motion, while the price in itself corresponds to the geometric Brownian motion.

The variance of the Wiener process, with or without a drift, is given by

$$\langle S^2(t)\rangle - \langle S(t)\rangle^2 = \sigma^2(t - t_0) \tag{6}$$

that is the distribution becomes wider with time, as a linear function of the time span.

A process with such proportional variance behaviour is called diffusive. In a different notation we can write

$$\mathcal{D}(l) = Dl \tag{7}$$

Where l is the time span $(t - t_0)$ and $D = \sigma^2$ is known as the *diffusion constant.* σ is often called *volatility* in financial applications.

2.1.1 Correspondence between the price difference and the return

In analysis it is not uncommon to use logarithmic prices. When price changes are small, there is a correspondence between the the logarithmic price difference and the return. The log-price difference

$$\log S(t + \Delta t) - \log S(t) = \log \frac{S(t + \Delta t)}{S(t)} = \log \frac{S(t) + \Delta S(t)}{S(t)} = \log(1 + \frac{\Delta S(t)}{S(t)}) \tag{8}$$

that is the logaritm of 1 plus the return. This logarithm can be expanded for small returns, so that it is shown to be nearly equal to the return itself:

$$\log S(t + \Delta t) - \log S(t) = \log(1 + \frac{\Delta S(t)}{S(t)}) \approx \frac{\Delta S(t)}{S(t)}. \tag{9}$$

2.2 Moments and kurtosis

Essential properties of probability distributions are captured in their *moments.* For the distribution of a random variable X, the *first moment* is μ, the expectation (the mean) of the random variable X. Adjusting for this mean, the nth *central moment* is

$$\mu_n = E((X - \mu)^n) \tag{10}$$

The significance of the first four orders of moments is as follows [29]:

- The **first central moment** is of course zero.

- The **second central moment** is the variance of the random variable, which measures how dispersed the possible values are around the mean, as is well known from fundamental statistics. As explained above (equation 6), the variance in a financial time series grows linearly with the time horizon of interest.

- The **third central moment** represents the lopsidedness of the random variable; this will be zero for any symmetric probability distribution. Dividing by the cube of the standard deviation gives the *normalized third central moment*, also known as the *skewness.* Although not taken into account in this thesis, Bouchaud et al. [7] report on a slight skewness of price series.

- The **fourth central moment** of a normal distribution is $3\sigma^4$. Normalizing (by division of σ^4) and subtracting 3, results in the *kurtosis*. It indicates the peakedness of a distribution, and is defined so that the kurtosis of a normal distribution will be zero. A positive kurtosis means a sharper and higher peak than the normal distribution, and also fatter tails. The variance of such *leptokurtic* distributions is more due to infrequent extreme deviations than to frequent moderate deviations. The leptokurtic distributions are of much interest in the analysis of financial time series, since price changes are not normally distributed, but have a positive kurtosis. This is disturbing for investors, because the fat tails of these distributions means that extreme occurrences are actually rather frequent, making investment fraught with risk of losing money.

3 Problem statement

3.1 Purpose

3.1.1 The purpose of agent-based modelling

Axelrod and Tesfatison [1] describe simulation in general – and agent-based modelling in particular – as a third way of doing science, in addition to deduction and induction. *Deduction* is deriving theorems from assumptions, *induction* is finding patterns in empirical data. *Simulation* produces data suitable for analysis by induction, but at the same time it allows the scientist to, in a way similar to deduction, start from a specified set of assumptions, instead of having to deal with the full complexity of real data from the studied social system.

Researchers in the agent-based modelling field are aiming at different specific goals. Axelrod and Tesfatison distinguish four of these: empirical, normative, heuristic and methodological goals:

1. Agent-based models can provide **empirical** understanding to why certain macroscopic features of a system, and not others, have evolved and persisted without much top-down control (e.g. standing ovations are not controlled by anyone in particular but the result of many individual decisions; certain bank receipts have through social acceptance turned into what we call money, others have not).

2. In **normative** investigations, the researcher uses the model as a laboratory for testing the qualities of different designs, looking for the one that gives the most desirable system performance over time (e.g. auction systems, voting rules).

3. The **heuristic** goal is the attempt to attain greater insight into the fundamental causal mechanisms in social systems. There is not always any obvious relationship between the simple assumptions used in the creation of a social system and the result that emerges from many agent interactions. For example, it may seem surprising to many that the fairy predictable behaviour of individual traders on a microscopic level can result in a virtually unpredictable price development on a macroscopic level. *This thesis is mainly concerned with the heuristic goal, since it investigates how the existence of different trading strategies affects the statistical properties of the emerging price curve.*

4. Finally, since agent-based modelling is a young field of research, the goal of **methodological** advancement is still an urgent one. The rigorous study of social systems through controlled computer experiments requires both careful consideration of methodological principles and practical development of the tools needed for implementing and evaluating the models. *In this thesis, the implementation has basically been made from scratch, and some fundamental methodological issues are considered in this report.*

3.1.2 The specific goals of this thesis

The thesis proposes an example of how an artificial double-auction stock market can be implemented, using an agent-based approach. It also briefly studies how different trading strategies affect the emerging price curve at an artificial market.

A first goal is to **create a functioning artificial stock market**, capable of evolving by itself once its initial state has been specified. Some of the basic requirements are:

- Output should be generated in the form of time series, corresponding to the trades and quotes data available from real markets.
- The emerging price curve should be "random walk-like", just like price curves in real markets
- The market should not die out within the time scale considered in the data analysis (i.e. the market should be in a state of *equilibrium*, see [25] for more on this matter)

A fundamental attitude has been not to tamper too much with the actual mechanisms of the continuous trading at real financial markets. For example, a continuous trading simulation should preferably not collect orders from several traders, in a round-based fashion, before matching orders and executing trades, but rather make sure that every new order immediately either is matched with a corresponding order or inserted into the order book.

At the same time, there are reasons to restrict the complexity of the model. With too realistic a model, the implementation would be very time-consuming. Also, a higher system complexity blurs the causal inferences.

In short, the design of the artificial market cannot be too simple, since it then would not be sufficiently similar to real stock markets for trustworthy results, but at the same time, it cannot be too complex, since it then would be more difficult to draw conclusions about the impact of specific factors.

Much of this report concerns methodological issues, and the implementation is carefully chosen to be as transparent as possible.

Going a step further, an artificial market could be used as a platform to **explore how different trading strategies affect the macroscopic behaviour of the market**. The goal here is twofold:

1. To start with extremely simple traders with the expectation that the causal mechanisms will be easily understood. These scenarios were used to briefly test the implementation's ability to generate robust results with reasonable outcomes.
2. Then, as a more specific task, strategies were developed to recreate the scenario imagined by Bouchaud et al. (as described in section 1.4.1 above) where the diffusiveness of price changes is a result of a fine-tuned competition between liquidity providers and liquidity takers. Although successful realizations would not prove any specific causal mechanisms, they would stand as an encouraging support for the logic presented in [7]. As a next step in this line of thought, one may be tempted to conclude that real traders actually think the way that the artificial agents do in the model. This, however, would be to draw the reasoning too far - other, completely different trading strategies may very well produce similar results and there is no way of telling by means of simulation which assumptions are correct.

As a final point, the greatest purpose of any research project is perhaps its potential to **inspire further research**. There are a number of ways to obtain this quality:

- *Results* can serve as a starting point for other projects or be compared to the results of others.
- *Tools* developed can be reused later on.
- New ideas and *creative reasoning* expressed by the researchers may inspire either themselves or others to pursue further projects in similar areas.

This thesis attempts to provide input to future research projects in all of the above ways.

3.2 Scope

Bearing in mind that there exists very elaborate artificial markets such as the Santa Fe Stock Market, which has been constantly improved by numerous post-docs, the scope of a master's thesis must be a modest one. Although it is a reasonable task to develop an artificial market whose basic functioning is similar to what happens in real markets, there is no possibility of mimicking all relevant aspects in a thesis such as this one.

The idea was to stay close to the analysis of empirical market data presented by Bouchaud et al. in [7]. Most of the statistical measures used in this thesis occur also in their article, and their ideas of the influence of different populations of traders are investigated.

The reader should keep in mind that many important aspects of trading activities have been excluded from the model. For example, everyone buys or sells only one share at a time, although trade volume certainly is an important factor in real markets (see section 7.5.2 for a further discussion).

4 Modelling and simulation issues

4.1 Implementation

4.1.1 The Java programming language

The artificial market was implemented as a Java application for several reasons:

- **Java is an object-oriented language**. Object-oriented languages in general lend themselves to agent-based modelling, since each agent can be described by its own object.

- **Java is widely used and supported**. The open source community provides ever-improving free Java libraries for most common development needs. The application of this thesis uses the *JFreeChart* package for the graphical interface and the *Colt Distribution* (created for CERN) for statistical analysis.

- **Java is free and easily shared**. A master's thesis cannot in itself give more than a modest contribution to the research in a given field, and every support for sharing the model is an asset. The Java language was designed for platform independency and using free software, applications can be compiled and executed on virtually any computer. Precise coding conventions and automatic API generation (javadoc) also facilitates the task for any researcher or student who wishes to pursue similar work.

- **Java is computationally efficient**. Although execution time never became an issue, Java is powerful in this matter.

4.1.2 General structure of the application

At the core of the application is an order book, where incoming limit orders are gathered and trades are cleared whenever an existing limit order can be matched with an incoming market order. The orders are sorted according to price-time priority, as described in section 1.3.2.

Connected to this order book is a number of agents (i.e. traders). Each agent is represented by it's own object in the implementation, so that they all can have different shareholdings, different trading horizons (i.e. on what time span they are investing) and even different rule sets for trading decisions. The whole population of agents can of course be *homogenic*, so that they all use the same trading strategies, but just as well *heterogenic* so that different subpolulations use different strategies. Most of the time, one single kind of traders or just two different kinds of traders were used in simulations.

Each type of agent has to be created directly in Java code, but once the needed set of agents is designed, the rest of the simulation can be handeled through a graphical user interface (GUI). The researcher controls the initial settings, chooses the simulation time interval and then lets the scenario evolve by itself. Some results are analysed by the application and presented in charts directly on the screen, and all important data written to files.

Parameter	*Typical value*	*Comment*
Number of agents	300	Important parameter, must be over a critical value for emerging properties to evolve.
Number of outstanding shares	900	Not important if short selling without limitations or fees are allowed.
Initial price	100.00 €	Relevant e.g. if price offers are relative to the current price.
Tick size	0.01 €	A small tick market
Short selling fee	0%	No fee in this implementation

Table 1: Some parameters on the market level.

Parameter	*Typical value*	*Comment*
Tendency to trade	0.01 €	Controls how often the trader will place a limit order.
Pricing variation	0.10 €	Controls how far from the reference price the trader will choose the price for her limit order (this will be the variance of some distribution).
Initial number of shares	3	Not important if short selling without limitations or fees are allowed.
Money to spend	100.00 €	Not important if money can be borrowed without delays or fees.

Table 2: Some parameters of a liquidity provider.

4.2 Parameters

The parameters of the model controls its initial state and also the market's behaviour during simulations. Just to give an example, some of the parameters at different levels are shown in tables 1 and 2.

4.2.1 Balancing the number of parameters

The *number of parameters* is one of the factors that determine how well a model can be fitted to empirical data. Although a high number of parameters makes it easier to get correspondence between simulations and data from real financial markets, there are good reasons to keep this number as low as possible. First of all, there is a theoretical issue: with too many parameters, there is a risk that so much flexibility is built into the model that it can be calibrated to fit well with empirical data whether or not it is a good model.

Moreover, exploring parameter range may reveal itself a very time consuming task if the parameter space has too many dimensions. When modelling a complex system such as financial markets, the parameter space easily grows into an unwieldy set of dimensions. Each feature added to the model often implies introducing several new parameters as well. In order not to get lost and sacrifice too much time on exploring the range of parameters that are less likely to impact the specific results that the researcher is after, a good philosophy might be simplify whatever can be simplified. By reducing parameters, it is easier to zero in on the key dimensions of the problem.

4.2.2 Reducing parameter space

Sometimes two parameters can be reduced to one, without simplifying too much. For example, the parameters

- "the *market attention* of an agent will control how often it surveys the market development", and

- "the *propensity to trade* of an agent will control how often it places or cancels orders when it considers that the market offers a good opportunity"

can be reduced to one single parameter:

- "the *likeliness to trade* of an agent will control how often it places or cancels orders when it considers that the market offers a good opportunity"

The idea is to let every agents pay as much attention to the market as all the others, but to instead adjust for how attentive they are in the *propensity to trade* parameter, which then becomes the *likeliness to trade.* Note that although the formulations of *propensity to trade* and *likeliness to trade* are seemingly identical, the latter will be lower for agents who are less attentive than others (*ceteris paribus*), while the former will not be altered because of agent's different attentiveness.

There is of course a slight difference between the two-parameter formulation and the one-parameter one when relating them to reality. It seems more realistic that some traders are more attentive than others, which would make it easier for them to spot good opportunities at the market. However, all such subtleties cannot always be captured in a model of reasonable size.

Many parameters have for practical reasons been set to fixed values, that sometimes correspond well to real markets, at other times differ from their real values, either because of lack of information or because of implementation issues. For example, agent rule sets would be more complicated if short selling came at a cost (which would be the realistic alternative) rather than if shares could be borrowed for free (which was the simpler alternative chosen for this implementation).

4.3 Simplifications

Although the artificial market of this thesis mimics many of the features of real stock exchanges, several simplifications were deliberately made to make the model easier to understand and less laborious to implement.

4.3.1 A one-stock market

The artificial market deals with one single asset. This is a very common simplification in artificial markets, and it is reasonable to assume that typical behaviour of financial markets occur also in single-asset markets. One would of course expect the demand for a certain asset to be altered somewhat if the option of trading other assets as well was added, since traders then would have other options for investment.

In any case, a basic understanding of a one-asset market seems like an important prerequisite for understanding what happens in more complex scenarios.

4.3.2 One share per trade

The perhaps most significant simplification consists in not allowing other order quantities than one share per trade. It is also a common simplification, but a more questionable one. Volume does play an important role in real markets, and it is not certain that market dynamics are not seriously affected by this simplification. However, this feature was left out in order to limit the implementation time. Especially, designing trading strategies would need much more attention if the agents would not only decide type of action (buy, sell or hold) and a price, but also a volume.

4.3.3 No external drift

No *external price drift* was imposed on the artificial market. Since the results were compared to *detrended* data (see section 5.1.2 below), the drift would not have been relevant. Furthermore, in high frequency data, the overall trend is generally small compared to intra-day price fluctuations.

4.3.4 Infrequent order cancellation

In real markets, an order is active in the order book until it has generated a trade or been cancelled. The default duration may be the rest of the trading day (currently used by e.g. Euronext, who operates the Paris Bourse), but traders may also cancel their orders whenever they feel opportune to do so. Judging from the statistics on the Paris Bourse presented in [3], 20-25% of limit orders are actively cancelled by the trader. In the artificial market of this thesis, orders were cancelled only when traders wanted to place a different order than the one they already had in the order book.

4.3.5 Intra-day effects only

No special treatment was used to account for overnight effects; during the hours when a stock market is closed, external events might still affect the opinions of the traders and, as a consequence, also affect the orders that will be placed when the market opens again. In the simulations, the total number of cleared trades was kept within the same order of magnitude as the number of trades per day for real, heavily traded stocks, and the continuous trading was assumed not to be interrupted during the simulated time interval.

4.4 Inherent constraints

As pointed out by LeBaron in [19], an elaborate set of agents may be indistinguishable from *zero-intelligence traders* (such traders, who always place their orders in the same price interval according to a uniform distribution, are investigated in the milestone article of Gode and Sunder [16]) when subject to the constraints of a market. An example of a crucially constraining rule quoted in the article is "never offer a lower price when selling an asset than what the cost was to obtain it". LeBaron draws the conclusion that researchers must be cautious about which features of the market are due to the intelligence of the agents and which are coming from the very structure of the market itself.

Some of the constraints tried out in my own market are easily turned on or off: allowing short selling removes the constraint on agents of only selling what they actually own; giving each agent limited funds imposes constraints on when to buy; dictating a maximum number of held shares in a single company is another way of limiting how much agents can buy, etcetera.

Other constraints are inherent in the model and were never turned off. Some of these constraints are in fact due to trading mechanisms and independent of the chosen implementation. The functioning of the order book makes sure that a trader who wants to buy does not trade with any seller, but only with the one offering the best price (it would of course be interesting to relax constraints such as this one, both for normative and descriptive purposes). As will be argued later on, this particular constraint has important consequences for the price fluctuations.

4.5 Time and correlation issues

4.5.1 Time discretization

The *quotes* and *trades* from double-auction continuous trading, which were used by Bouchaud et al. [7], were labelled with time stamps accurate to the second. Although this is *time-discrete* data, the underlying processes are not - traders take their actions and make their decisions in the real world where time is *continuous*.

Thus, when creating an artificial market on a computer, the consequences of the transition from continuous time to discrete time have to be taken into consideration.

On a more concrete level, a discrete time requires that an order of actions must be imposed. For example, if all the traders on a market should be able to act at time t before proceeding to time $t+1$, several traders might want to buy or sell at the same time instance t – but who will have priority? One way of dealing with this issue is to go through all the traders one by one, letting each have a look at the order book and any other information available at the time, make her trading decisions, place or cancel orders and letting the market handle any trade before proceeding to the next one. After the round is completed, it can start all over again.

However, if the traders get to act in the same order every round, they will each time be looking at the results of the actions of their closest neighbour. Some positions in this chain of neighbours will probably be more advantageous than others[1], and in any case, *correlations* will be introduced between the actions of different traders. These correlations are of an unwanted kind, since they are not the direct result of an attempt to mimic real behaviour, but due to fundamental limitations in the model – although certain traders perhaps do look closely at what other traders are doing, most of them will probably not scrutinize the order book right after the actions of the same competitor all the time, at least not in a way that links all traders in a chain of observation.

4.5.2 Two kinds of correlations

Two kinds of correlations exist in the model:

1. First, there are correlations that may **exist in the real world**: for example, a trader might base her decision on an average of the price over the past few hours (correlation in time) or upon the action of another trader that she believes is informed (correlation between agents). These are important features of the market, and if such correlations are correctly observed and then mimicked in the simulation, more of the rich characteristics of financial data can be explained by the model (e.g. possibly the fat tails of the return distribution). Of course, if false assumptions are made about the mechanisms of the market, this will result in the wrong correlations, so it is still important to be restrictive enough to include only the aspects necessary to render data sharing the characteristics of real-life data.

2. Second, there are correlations that cannot exist in real-life markets, but which are **due to modelling imperfection**. As mentioned in the previous section, if the trading is modelled in rounds with all traders traversed in the same order at each round, trader i will always consider the order book right after trader $(i-1)$, so she will never know what it looks like directly after trader $(i-2)$ has placed her orders. This sort of correlation is purely due to the imperfections that any model suffers from, and their impact must be assessed and reduced if possible.

4.5.3 Reduction of unwanted correlations

As the looking-at-one's-neighbour correlations are a consequence of the suggested sequential computer implementation, it might be transformed or disappear with a different approach. A desirable solution would be to let the order book and all of the agents each be running in its own thread, letting all threads run simultaneously in the simulation (there exist pre-programmed platforms with the purpose of serving as tools for agent-based modelling [21]; some of which may very well be designed to support multi-thread execution.) Although the Java language offers support for multiple threads, this approach would have required more programming skills than mastered by the author of this report,

[1] Compare the situation to an ordinary game of poker: The betting goes in a clockwise manner, which means that you will know whether the person sitting to your right is calling, betting or folding, while you'll have to anticipate the actions of your neighbour on your left. If a certain player has an unpredictable playing style, you would probably prefer having this person seated to your right, so that you already have information about her actions when you're in turn. A predictable player, on the other hand, could just as well be seated on your left, since you have a good chance to anticipate her actions anyway.

and a more modest choice was made: to randomize the order of the agents in each round.

A different solution that was also considered, but rejected as unnecessarily consuming in terms of computational power, was lowering the probabilities that an agent would want to place an order, so that all of the traders would have been traversed many times before the next trading decision was made. The effect of looking-at-your-neighbour would thus have been diluted, at the cost of slow execution.

4.5.4 Time notions in the model

In their analysis, Bouchaud et al. [7] use *trade time* (not to be confused with *trading time*, see appendix A), that is a discrete time notion where a new time step is taken at each trade, irrespective of how much (real) time actually has passed since the previous trade. In order to enable comparisons, trade time compability was the basic requirement on the time notion used in the model, with regard to data output.

In simulation, a finer time grid was obviously used. In order to extend analysis to also encompass volatility issues, a time notion that somehow corresponds to real time would be useful. For this purpose, the artificial market was also designed to record *agent time*, defined as a time notion where a time step is added each time another agent is in turn to act. The idea was to let an agent time step correspond to a real time interval, say a millisecond. In periods where the price was changing more quickly, agents could be designed to act with a higher probability, which would result in fewer agent time steps, that is less real time between each trade – a realistic behaviour with respect to volatility.

4.6 Decision making

An intricate part of creating an artificial market is to design the rules that will let agents decide their trading action (buy, sell or hold) and, in the case where they decide to place an order, the price and the quantity of the offer. Some of the difficult aspects are:

- **Information availability.** Which information apart from the current order book will be available to a trader making a decision? Will information be delayed? Will information be costly?

- **Attention and choice of sources.** Which of the information available will actually be used by traders? How far back in the price history do they look? How often do they look at the order book? Do they pay attention to the behaviour of other traders?

- **Trading strategies.** How does the information translate into actual trading decisions? Does one decision involve several trading actions spread out in time? What is the time horizon of the investments? Do traders continuously adapt their strategies to better fit the current trading situation?

Many artificial markets aim to investigate which information is actually used by traders. They often use evolutionary environments, where agents with a winning use of information survive, while the others perish. One of the most appealing features of such *adaptive models* is that the trading strategies are not imposed on the model by the researchers, but found *endogenously*. The winning strategies are then a result from the simulation, not something that the researchers had to guess beforehand and try out. The downside is that the logic behind these winning strategies stays hidden in the model.

This thesis uses the opposite approach. Trading strategies are explicitly formulated and implemented. There is a risk that these made-up strategies are unrealistic, since researchers cannot know what actually goes on in the brain of a trader (even the traders themselves are perhaps not always aware of which factors are the most decisive when trading decisions are made). On the other hand, imposed strategies can be kept simple but still representative for very typical behaviour. The researcher can verify if a well-defined trading strategy has a certain impact, in a robust manner, on the statistical properties of the emerging price curve. In this thesis, the strategies were designed as simple as possible (see section 4.2.2), and although all agents were both selling and buying, some agents acted solely as liquidity providers, others solely as liquidity takers.

4.6.1 Price notions for trading decisions

The agents of the artificial market generally based their decisions on the current price only, and did not look at past prices. However, the "current price" is not an unambiguous term. Looking at the order book, a trader could consider the best bid, the best ask or perhaps the midpoint of the spread. She could also look at the price of the last trade that was actually executed. Depending on what the trader wants to do, the preferred price notion to use will probably vary. When creating a market buy order, it seems reasonable to consider the current best ask. When creating a limit buy order, it would probably be wiser to choose a price close to the current best bid. Different price notions are listed in table 3.

4.7 Testing

4.7.1 Programming errors

There are several ways of testing for programming errors. Special concern was made to check the Java methods dealing with the statistical treatment of the artificial market data. This testing included

- applying implemented algorithms on a pure random walk (where it is clear from theory what the result should be) before applying it on the price series,
- running parallel data analysis in Matlab, and compare the respective results testing the algorithms on markets with predictable populations, using extremely simple rule sets,
- comparing the results with the analysis by Bouchaud et al. in [7] and blame any discrepancy on programming errors in the first place, before drawing other conclusions.

Price notion	*Notation*	*Definition*	*Comment*
Best bid	b	The limit price of the current best buy offer in the order book.	At the lower side of the spread.
Best ask	a	The limit price of the current best sell offer in the order book.	At the upper side of the spread.
Midpoint	m	The average of the best bid and the best ask.	In the center of the spread.
Traded price	*(not used)*	The price of the last trade.	Shows what someone actually paid, but ignores all changes in the order book since the last trade.

Table 3: Different price notions for trading decisions.

4.7.2 Checking the model for parameter dependence

The behaviour of artificial markets can sometimes be very sensitive to parameter values. In the worst cases, the whole functioning of the market depends on the fine-tuning of one critical parameter: see for example [20] on how the Santa Fe Artificial Stock Market was very sensitive to a parameter that adjusted the price to reflect excess demand or supply.

In other cases, different intervals of a parameter setting may all render consistent results, but at the same time reveal different market behaviour: when agents at the Santa Fe market seldom updated their trading rule sets, they tended to base their decisions solely on rational expectations of the future share price; however, when learning more quickly, they began to use other sources of information as well, including technical trading information [19].

In a setting when parameter space is reduced as much as possible, it is still a good idea to run different realizations while altering fixed parameters one by one. This is a way of checking for the type of extreme parameter sensitivity that may be of decisive importance for the robustness of a market. Especially during the development phase of the market, when new functionalities are added gradually, conclusions about the impact of each added feature should not be drawn before some of this basic testing has been carried out.

Price notion	*Notation*	*Definition*	*Comment*
Traded price	*(not used)*	The price of the last trade.	This price is determined by a limit order from the upper part of the spread if the trade was triggered by a market buy order, but from the lower side of the spread if the trade was triggered by a market sell order.
Preceding midpoint	p_n	The average of the bid and ask of the quoute that immidiatley preceded trade n.	In the center of the spread of the preceding quote. Fluctuates less than the traded price, since it is independent of the sign of the trade.

Table 4: Different price notions for analysing trade data (cf. the price notions for trading decisions in table 3)

5 Data output and data analysis

5.1 Conventions

The statistical analysis stays very close to Bouchaud et al. [7], using the same concepts and notation. Therefore, the data output from the artificial market was structured in the same way as the data they were using: one series of all successive trades and one series of all successive quotes.

5.1.1 Price notions for data analysis

At each instant in a double auction one-stock market, there are several price notions to consider, as mentioned in section 4.6.1. The price for anyone who wants to buy at market is the best ask, a, while anyone who wishes to sell at market will have to do so at the best bid, b, which is lower.

When analysing trade data, the *traded price* is perhaps the most natural way of defining the price of any specific trade. However, since the price of a trade is determined by the limit order involved, the traded price will be taken from the upper side of the spread if an incoming market buy order is matched with a limit sell order in the order book, but from the lower side of the spread if the incoming market order is a sell order. The curve of the traded price could thus shift a spread-width up or down, depending on the sign of a trade, even if the actual value of the stock were constant.

A more stable price notion is to consider the midpoint $m = (a + b)/2$ preceding each trade, which is also the choice of Bouchaud et al. [7] For all data analysis presented in this report, the price is defined as the midpoint preceding each trade.

Different price notions for data analysis are shown in table 3.

5.1.2 Detrended data

When looking at financial data, there is generally a *long-term upward trend* in prices. It can fluctuate, though, and when the analysis is isolated to a specific business the trend can just as well be downward. Especially in *low-frequency data*, not only the performance of a company determines the price development of its stock, but also the success of the business it belongs to. In *high-frequency data*, trends tend to disappear in the fluctuations.

If some sort of overall price development is known, data can be detrended, that is adjusted in a way that compensates for the overall drift. This is often done in analysis, and was done in [7], the reference point of the data analysis of this thesis. No external drift was added in the simulations accounted for in this report, and no detrending of the output data was needed to make valid comparisons.

5.2 Observables

5.2.1 Return autocorrelation

The *return* on an investment at time n is the profit relative to the investment:

$$r_n = \frac{p_{n+1} - p_n}{p_n} \tag{11}$$

As a first indicator of random walk-like behaviour of the price series, the *return autocorrelation* was analysed:

$$\mathcal{C}_r(l) = \frac{\langle r_n r_{n+l} \rangle}{\langle r_n r_n \rangle} \tag{12}$$

where the denominator is a norming factor, making sure that the value of the function will be 1 for $l = 0$. Unless returns are correlated in time, this autocorrelation function will be zero for all other l-values.

5.2.2 Distribution of returns

The *distribution of returns*, that is the distribution whose cumulative distribution function is defined by

$$F(x) = P(\frac{p_{n+1} - p_n}{p_n} \leq x), \tag{13}$$

is expected to be Gaussian in settings without correlations (see section 2.1). In real markets, however, the distribution is known to be leptokurtic, with fat tails. The distribution was constructed through a histogram, where the bin-size was chosen narrow enough to clearly see kurtosis effects, but wide enough to give a smooth distribution.

5.2.3 Standard deviational constant

Bouchaud et al. [7] considers the *average mean square fluctuation of the price* between time n and $n{+}l$:

$$\mathcal{D}(l) = \langle (p_{n+l} - p_n)^2 \rangle \tag{14}$$

According to the diffusiveness of price changes, the following relation should hold

$$\mathcal{D}(l) = Dl \tag{15}$$

where $D = \sigma^2$ is the diffusion constant. A way to study the behaviour of $\mathcal{D}$ is to plot $\sqrt{\mathcal{D}(l)/l}$, which should then be constant, dimensionless and equal the standard deviation, σ. Due to its expected properties, this entity could be called the *standard deviational constant*, although it will only be constant if the price development really is perfectly diffusive.

This function plays an important role in the thesis. Since one would expect it to be decreasing for a sub-diffusive market and increasing for a super-diffusive market, it gives some indication of the overall diffusiveness of the model.

5.2.4 Sign autocorrelation

Although empirical studies reveal little correlation in price changes, one finds – surprisingly enough – slowly decaying correlations in *trade signs*, ε. As mentioned in section 1.3.4, the sign of a trade is defined as

$$\varepsilon = \begin{cases} 1, & \text{if the trade was triggered by a market order to buy} \\ -1, & \text{if the trade was triggered by a market order to sell} \end{cases} \tag{16}$$

The *sign autocorrelation*, defined as

$$\mathcal{C}_0(l) = \langle \varepsilon_{n+l} \varepsilon_n \rangle \tag{17}$$

is considered in [7]. Since we know that the random-walk-like price changes are not correlated, it is surprising to see that the trades - who are the immediate source of price changes - show slowly decaying correlations with respect to their sign.

According to Bouchaud et al., the slowly decaying sign autocorrelation is easily explained by the strategies of liquidity takers. By dividing large orders into several smaller ones of the same sign and spread them out in time, they create correlated trades. The authors argue that there must be a compensating mechanism so that price changes remain uncorrelated, and they propose that this mechanism is induced by the strategies of liquidity providers.

5.2.5 Response function

Every trade potentially contains some information on the stock value. The impact of trading on price changes can be studied via the response function:

$$\mathcal{R}(l) = \langle (p_{n+l} - p_n) \cdot \varepsilon_n \rangle \tag{18}$$

This function measures how much the price moves up between time n and time $n+l$, conditioned to a buy order at time n (or moves down, conditioned to a sell order). Bouchaud et al. [7] carefully point out that, due to the correlated trade signs, this is not the market response to a single trade (in their article they also theoretically investigate the relationship between these two entities). The authors present an analysis of empirical data, where the response function is almost constant, at least within a factor two (they detect an initial slow raise followed by a decline after 1000 trades or so).

6 Simulation results

6.1 Different kinds of traders used: an overview

During simulations, different kinds of traders were used. Table 5 gives an overview of all the traders used, and which results each trader type has contributed to. The *Sections* column of the table shows all results sections in this report that refer to the specific trader type. The setion where the trader's strategy (and purpose) is presented more in detail is marked in bold face.

Name	*Description*	*Special parameters*	*Sections*
ZeroIntelligenceTrader	Creates orders at random (buy or sell at equal probabilities) in the interval]0,200[, a market order if a better offer already exists in the order book, a limit order otherwise.	*no special parameters*	**6.2**, 6.3
RandomTrader	Creates a limit order with a price taken from a normal distribution centered around the current best corresponding offer, but converts it to a market order if a better offer already exists in the order book.	σ_p^2, the variance of the order price	**6.4**, 6.5, 6.5.1, 6.5.2, 6.6, 6.7, 6.8, 6.9
EagerTrader	Takes any offer already existing in the order book, but when facing an empty order book she creates a limit order with a price taken from a normal distribution centered around the last traded price.	σ_p^2, the variance of the order price	**6.5.1**

Table 5, *continued on next page...*

Name	*Description*	*Special parameters*	*Sections*
LiquidityProvider	Creates a limit order with a price taken from a normal distribution centered around the current best corresponding offer, but refrains from creating an order if a market order would have been more advantageous.	σ_p^2, the variance of the order price; P_{skip}, the probability of skipping a turn (used for balance in two-population scenarios)	**6.9**, 6.10.1, 6.10.2
RandomInformedTrader	Places market orders of the same "sign" (buy or sell) only, at a certain probability.	λ, likeliness to place an order when in turn	**6.9**
SerialTrader	Creates a correlated series of trades of the same "sign" (buy or sell), and refuses to make a worse deal than the first one later on during the series. The series commences at a random time.	λ_{series}, likeliness to start a series of trades	**6.10.1**
ExpectingTrader	Creates a correlated series of trades of the same "sign" (buy or sell), and refuses to make a worse deal than the first one later on during the series. The series is more likely to start when there is a large gap between the current midpoint and the traders expected price.	λ^*_{series}, modified likeliness to start a series of trades	**6.10.2**

Table 5: (*... continued from previous page*) The different kinds of traders used in the simulations of this thesis. For a more detailed description of the trading rules used by a particular kind of trader, refer to the section marked in bold face.

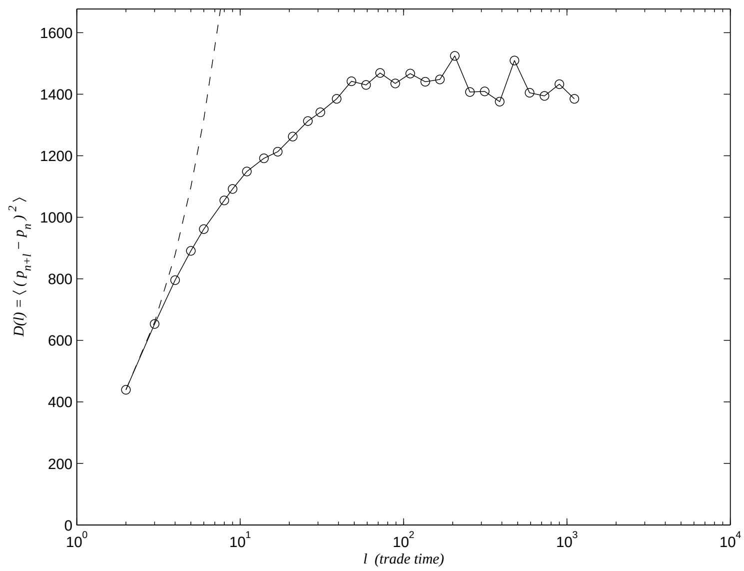

Figure 3: The average mean square fluctuation of the price in a realization with 300 *ZeroIntelligenceTraders*. The dashed curve shows how a linearly growing function would appear in the diagram. There is no evidence of diffusive behaviour.

6.2 Diffusiveness with zero intelligence traders

The market was run with a population of agents of the type ***ZeroIntelligence-Trader***, a trader who

1. places random limit orders (buy or sell with equal probabilities) in the price interval $]0, 200[$,
2. converts the limit order to a market order if a better (or equally good) deal than the limit price is already offered in the order book,
3. only puts in a new order if she does not already have one of the same "sign" (buy or sell) in the order book,
4. cancels an existing order if and only if she is about to emit a new order of the opposite "sign".

This type of trader does not pay any attention to the actual price of the shares, which ought to suppress the diffusiveness. The $\mathcal{D}$-function of equation 14 and its dimensionless version, the standard deviational constant, are plotted in figures 3 and 4. The variance clearly does not grow linearly with time; the process is sub-diffusive.

6.3 Distribution of returns with zero intelligence traders

A basic criterion for the artificial market was that the emerging price curve should be close to a random walk. More specifically, the distribution of returns

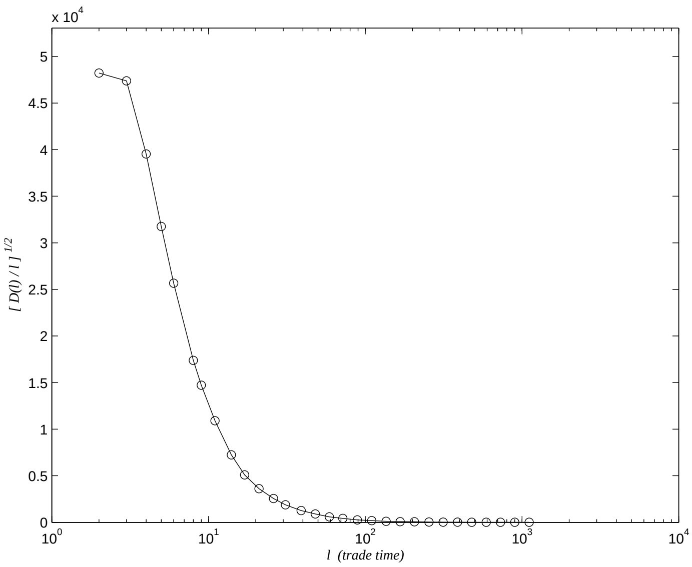

Figure 4: The heavily sloped curve of the standard deviational constant in a realization with 300 *ZeroIntelligenceTraders*.

should not be too far from Gaussian. From empirical data we know that the distribution is actually leptokurtic, that is with fatter tails than Gaussian.

Interestingly, zero intelligence traders are capable of reproducing several important features of financial markets (as also confirmed in e.g. [6]). In the simulations, the distribution of returns went leptokurtic as soon as the number of agents exceeded just a few individuals; in figure 5 this distribution is shown for a population of 300 agents for a realization over 10 000 trades. For most realizations, the distribution resembled this one, being fairly close to Gaussian, but leptokurtic.

The leptokurtosis is more easily seen if the vertical axis is log-scaled, and in the two diagrams of figure 6, this is shown for a realisation with 300 *ZeroIntelligenceTraders*. For clarity, a realization spanning over 100 000 trades provided the return values, since shorter realisations sometimes gave distributions that were slightly skewed.

The last two points of the *ZeroIntelligenceTrader* decision rules above was in one realization replaced by the alternative of always placing a new order, cancelling any existing own order, regardless of its “sign”. The results were very much the same.

6.4 A market in equilibrium

An attempt was made to create a basic type of trader that was not completely of zero intelligence, but actually has a look at the current price before putting in any orders.

The ***RandomTrader*** is a trader who

1. places random limit orders (buy or sell with equal probabilities), where

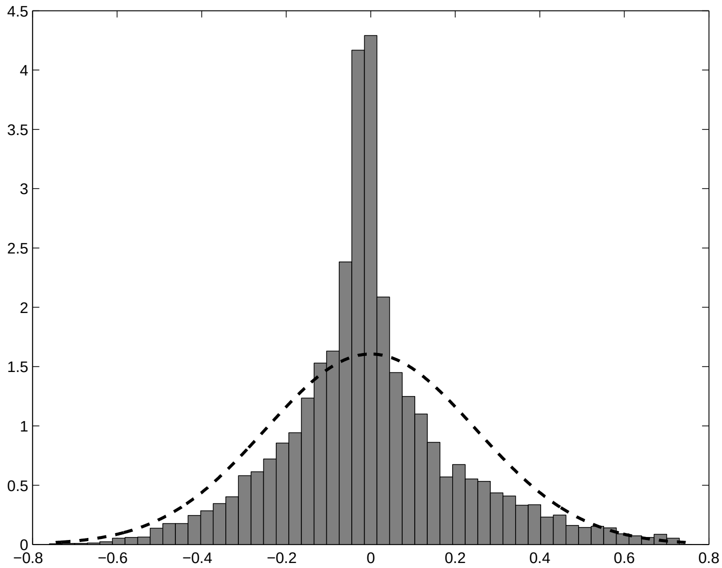

Figure 5: The leptokurtic distribution of returns for a realization with 300 *ZeroIntelligenceTraders*, over 10 000 trades. The histogram is created from the returns and the bell shaped curve of a normal distribution is added to the chart for comparison.

the price is taken from a normal distribution centered around the current ask if it is a sell order, around the current bid if it is a buy order,

2. converts the limit order to a market order if a better (or equally good) deal is already offered in the order book,
3. only puts in a new order if she does not already have one of the same "sign" (buy or sell) in the order book,
4. cancels an existing order if and only if she is about to emit a new order of the opposite "sign".

This trader was for example used to check whether the model was stable or not for simulating a certain period of *trade time*. In general, the realizations of this thesis encompassed 10 000 trades, which roughly corresponds to a trading day for the most frequently traded stocks (up to one trade per second [7]). In order to make sure that the artificial market did not die out within the time span of interest for analysis, realizations with different types of agents trading at random were run until 100 000 trades were cleared. In all cases, the price curve was still random walk-like and in the same order of magnitude as the initial price, so the result was deemed satisfactory. One of the realisations is shown in figure 7.

6.5 Impact of price notions in the analysis

As an illustration of how the choice of price notion in analysis affects the distribution of returns, data from the same realization was analysed using different

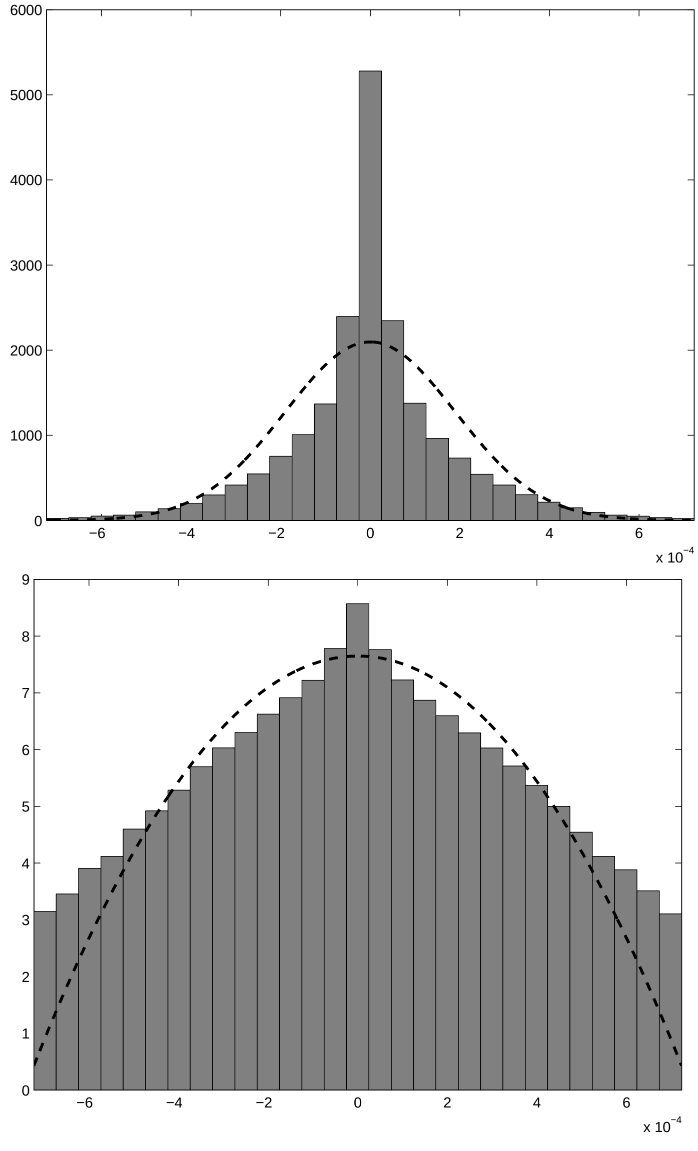

Figure 6: The leptokurtic distribution of returns for a realization with 300 *ZeroIntelligenceTraders*, over 100 000 trades. The upper diagram shows the distribution with linear axes, the lower one shows the same distribution with a logarithmic scale on the vertical axis.

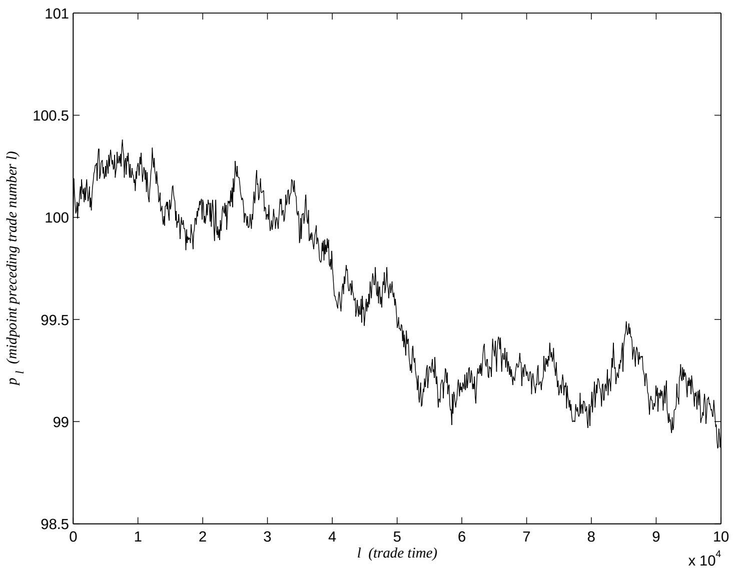

Figure 7: The price curve of a realization comprising 100 000 trades, showing that the trading activity does not collapse or die out even within a much larger time span than the one used for analysis. This was tried out for several different trader populations; in this case it was 300 *RandomTraders*
.

price notions in figures 8 and 9. The agent type used for this realization was a modified version of the *RandomTrader*. It chooses to place a market buy order instead of a limit buy order not only if a better deal already is offered in the order book, but also if the preferred price is close enough to an existing offer, where "close enough" is defined as "not more than half-way across the spread".

There is no intention here to go further into this matter, but is seems established that unless compensated for, the shifts due to trade signs may strongly affect the distribution of returns in this high frequency data. Preceding midpoints were used in this thesis.

6.5.1 Influence of order book mechanism on return distribution

As mentioned in section 6.2, the return distributions very easily turned leptokurtic. Actually, the only scenario where the distribution of price changes turned out to have a close-to-zero kurtosis was populated by ***EagerTraders***, agents so eager to trade that they accepted any offer that was in the order book, or placed an own offer at random (with at Gaussian distribution) if there was no pre-existing order. In other words, in this scenario the limit price-sorting function of the order book, that makes sure that the best offer is taken first, was not in use; any offer was accepted.

Since it seemingly was the order book mechanism that increased the kurtosis of the distribution, a scenario was designed to lie in between the one using *EagerTraders* (where the order book is barely in use) and the one using *RandomTraders* (where the order book is very much in use). This was realized

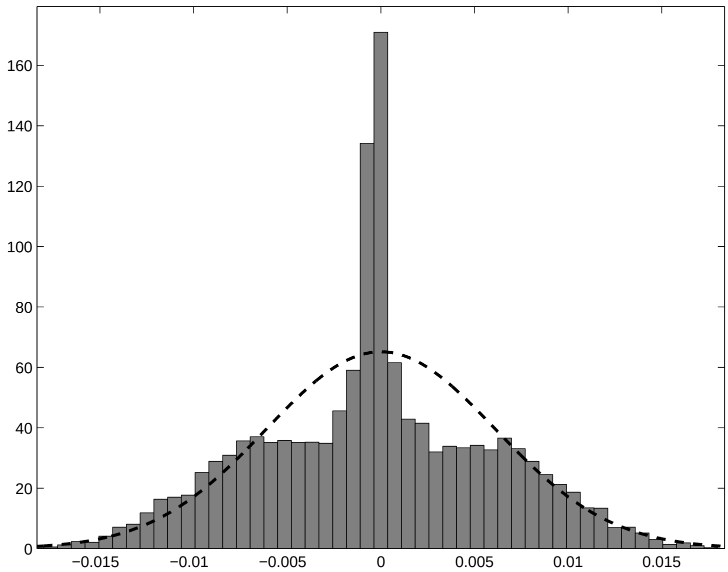

Figure 8: Using the traded price in analysis is less stable than the midpoint price. This is the deformed distribution of returns for a realization with 300 random traders that use market orders rather than placing limit orders across half the spread. The histogram is created from the returns (using the traded price) and the bell shaped curve of a normal distribution is added to the chart for comparison. With a higher (but unrealistic) variance for the order limit prices, the distribution splits up into three different peaks, separated approxiamtely by a spread-with. This tendency is visible in this histogram as well, although not as prononced, since the three peaks overlap.

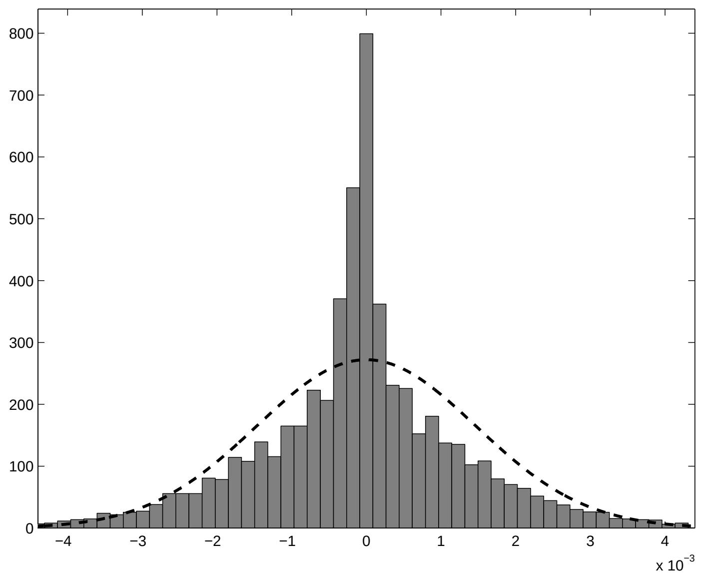

Figure 9: The distribution of returns for the same realization as in figure 8, but now calculated with the preceding midpoints as the price notion. Price changes seems more normally distributed when the "sign" of the trade has less influence on analysis.

using a polulation with *RandomTraders*, but where limit orders were cancelled after a specific time period. As long as the order duration is generous, many orders remain in the order book and the situation is close to a realization with ordinary *RandomTraders*. Cancelling with a short order duration, on the other hand, keeps the number of orders in the order book to a low value. As duration decreases, the scenarios approaches realizations with the *EagerTrader*.

Results are shown in figures 10 and 11. For most realizations in this thesis, the return distribution and the distribution of price differences look the same, and it is from a financial perspective more interesting to speak about returns. With *EagerTraders*, however, the price changes may be very large, and the two different distributions are no longer approximately equal. In this particular case, the price difference distribution (figure 10) is more easily interpreted.

6.5.2 A tick effect

When the bid-ask spread was large compared to the *tick* (the smallest price difference allowed) the histogram-generated distributions had the rather smooth look, as in figure 12. However, when the parameters of the trading strategies were modified to diminish the spread into more realistic proportions, the distribution was less regular, as shown in figure 13, where all parameters have the same values as in figure 12 except for the tick size.

The problem is that the tick limitation on price offers generates price changes in whole ticks only. With a fluctuating price, the return values could still, in principle, assume any value. If price fluctuations are small, however, the discretized price changes will all be divided by similar price values, so that the

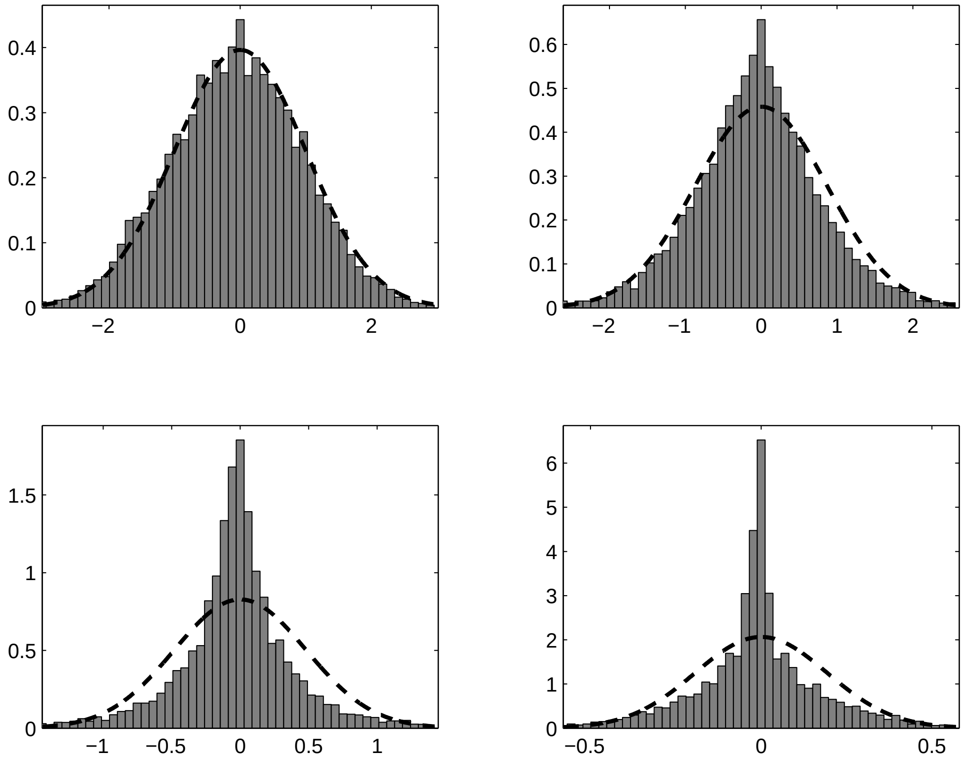

Figure 10: Four different **distributions of price differences** in realizations with 300 traders using a variance of 1.00 € when choosing order prices from a normal distribution centered around the last traded price (initiated at 100.00 €). The populations consisted of *EagerTraders* (upper left diagram), *RandomTraders* who cancelled their own orders as soon as one other trade had occurred (upper right), *RandomTraders* cancelling after four other trades (lower left) and *RandomTraders* without these special cancellations (lower right). The distribution for the *EagerTraders* is close to Gaussian, but it gets increasingly leptokurtic as the traders use the order book with more patience.

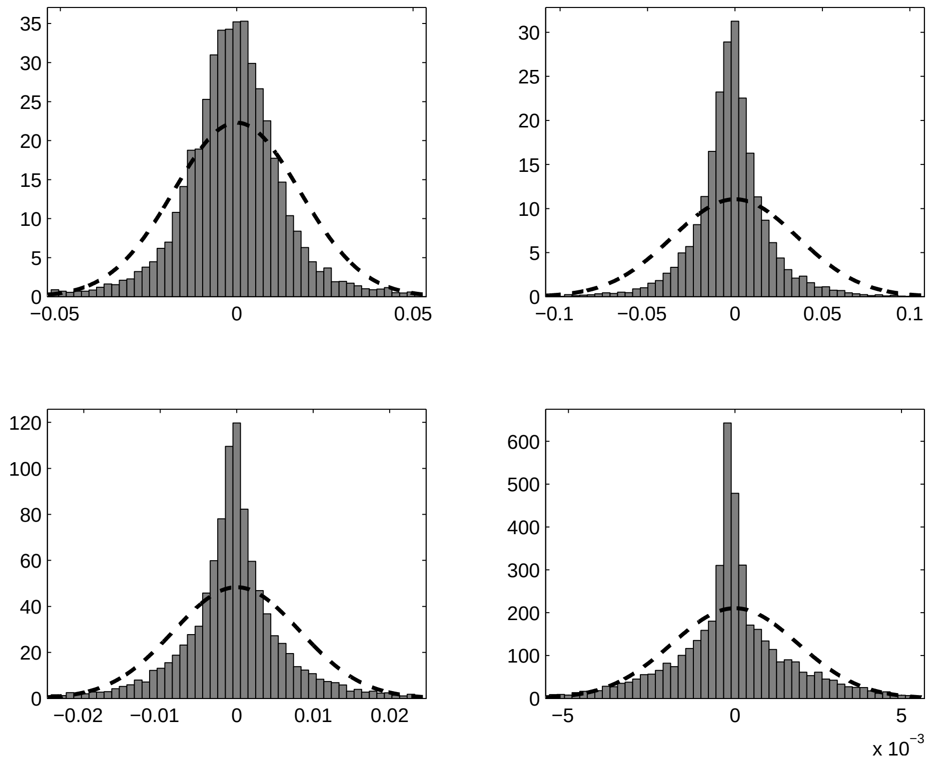

Figure 11: Four different **distributions of returns** in realizations with 300 traders using a variance of 1.00 € when choosing order prices from a normal distribution centered around the last traded price (initiated at 100.00 €). The populations consisted of *EagerTraders* (upper left diagram), *RandomTraders* who cancelled their own orders as soon as one other trade had occurred (upper right), *RandomTraders* cancelling after four other trades (lower left) and *RandomTraders* without these special cancellations (lower right). Unstable price changes makes these diagrams harder to interpret than their correspondents based on price differences, shown in figure 10.

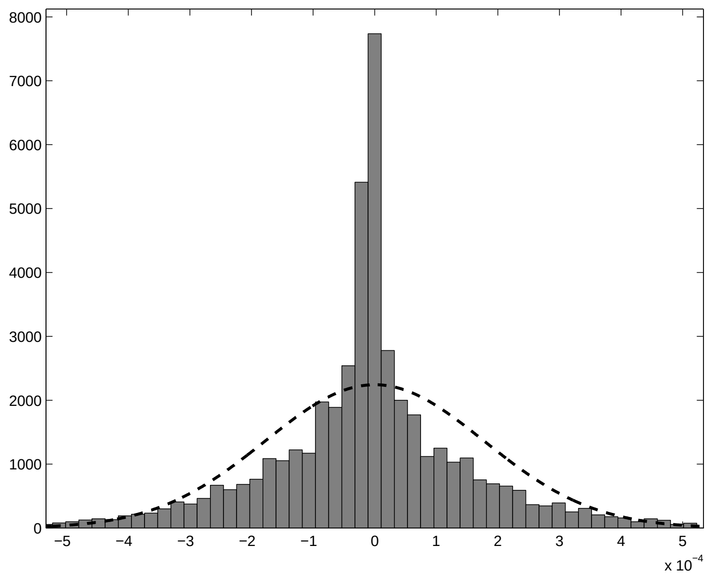

Figure 12: The **distribution of returns** for a realization with 300 *Random-Traders*, where the average spread was 0.05 and the tick size 0.001, perhaps an unrealistically low value. Since the tick size is much smaller than the spread, the distribution is rather smooth.

returns will not reach certain intervals, but instead be clustered into some of the bins. One simple solution is to use a lower number of bins, as shown in figure 14.

This tick effect did not seem to disturb other observables rendered by the simulation, and since the return distributions were not studied in detail, no further work was done on this.

6.6 Short selling and limited resources

A realization was made where *short selling* was removed from the model, so that traders making random decisions were not allowed to submit their sell orders unless they actually had some shares to offer for sale. The result was a distinct and linear upward trend in the price of the share. The mechanism is easily explained: if some of the traders that see a good opportunity for selling cannot place any orders, but every trader that believes buying would be profitable still can place orders, the demand will exceed the supply, causing a price raise according to common market logic.

A similar effect occurred in realisations where each agent had a limited amount of money to invest, so that it could not buy when its purse got empty. This rule gradually excluded more traders from buying (but not from selling) as the simulation went on, and prices collapsed, since there were too many sellers compared to the number of buyers.

These realizations (where *RandomTraders* were used) with a “no short selling” constraint and a “no buying unless affording it” constraint, respectively, probably says more about the model than about the real world. For example, in

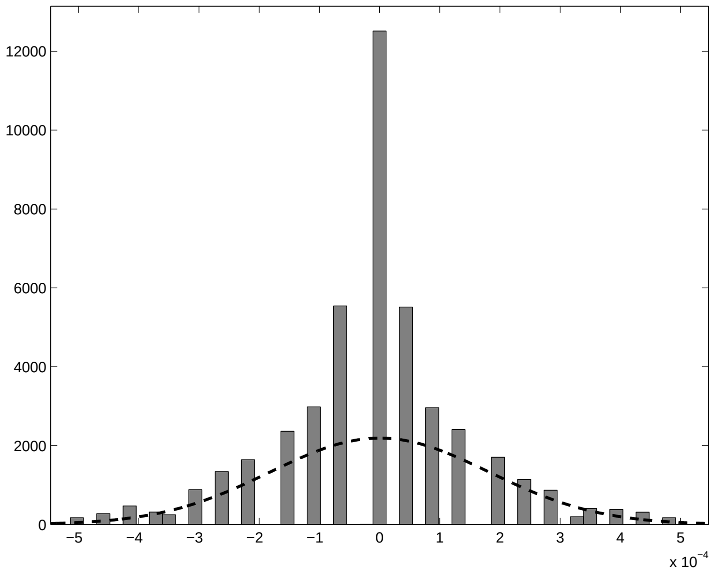

Figure 13: The **distribution of returns** for a realization with 300 *Random-Traders*, where the average spread was 0.05 and the tick size set to the realistic small-tick value of 0.01. With a tick size in the same order of magnitude as the average spread, there are gaps in the distribution.

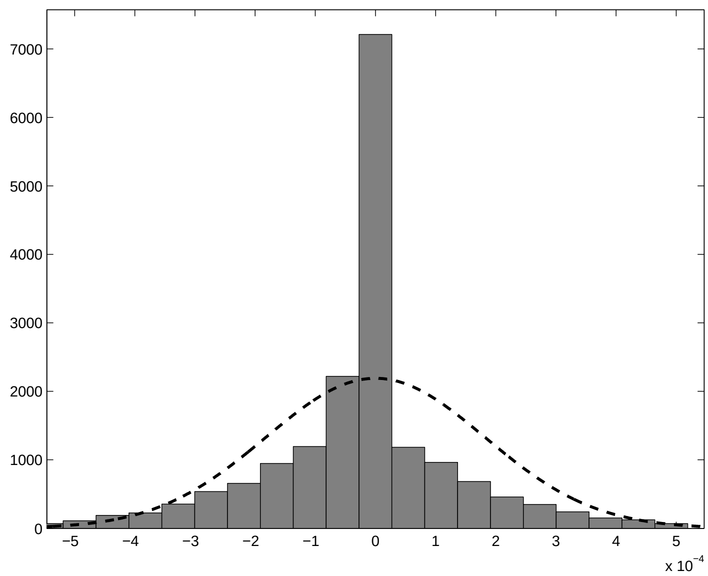

Figure 14: The **distribution of returns** for the same realization as in figure 13, with wider bins that coarsly cover the gaps.

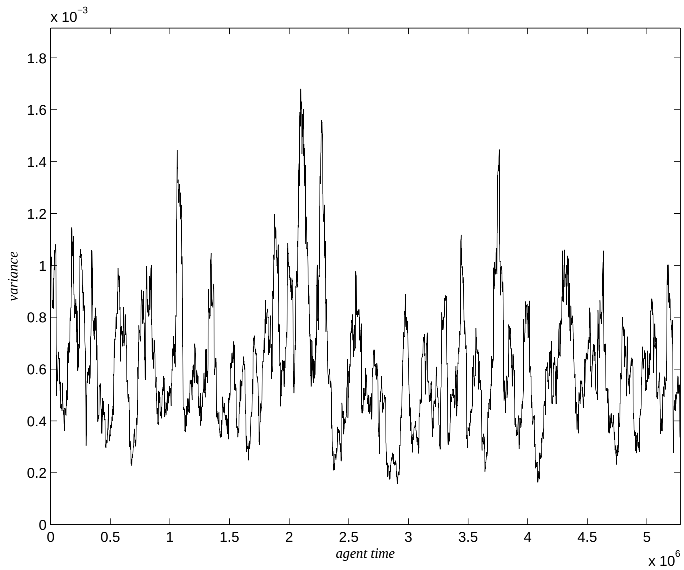

Figure 15: The fluctuating **variance of price differences**, from a realization of 300 *RandomTraders*, using a moving window of 50000 *agent time* steps.

the case of no short selling, real traders may be sensitive to *absolute prices* (e.g. they might believe that there is a fundamental price), which would limit the price raise, or there might be important share owners that adjust their supply to the overall demand, which would eliminate the demand surplus that caused the price raise in the realization. Possibly, the result indicates that if a real market deals with a single asset that cannot be remanufactured and where no short selling is allowed, there will be constant inflation in prices. If a slight effect of this kind exists in stock exchanges as well, it will probably disappear, though, in the stronger upward trend that is due to the arbitrage relationship between the stock exchange and a fixed rate interest market.

Since the model turned out to be sensitive (cf. section 4.7.2 on parameter dependence) to imposed constraints that skew the balance between *bulls* and *bears* (see appendix A for explanations of these terms), in all other realizations short selling was allowed and every investment considered to be affordable for the traders.

6.7 Variance

The notion of *agent time*, introduced in section 4.5.4, was used to calculate a variance of price changes. The price was sampled at equal time distances in *agent time*, which means that a different number of trades could have occurred between each sample. The sampled data points were then used to calculate the variance with a moving time window.

In figure 15 this variance is plotted for a realization with *RandomTraders*, using a window of 50000 *agent time* steps, roughly corresponding to 100 trades on average. As in real markets, the volatility is fluctuating.

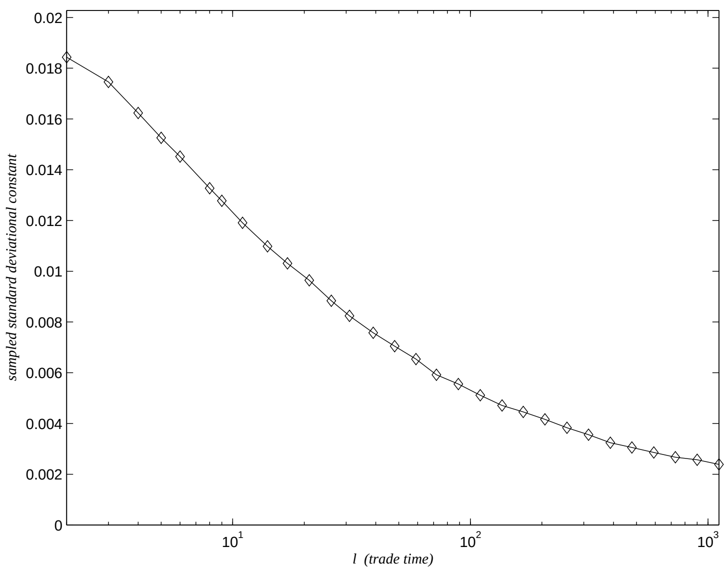

Figure 16: The *agent time* version of the **standard deviational constant**, from a realization of 300 *RandomTraders*, using a sample frequency of 1000 *agent time* steps.

The same time notion (and the same realization) was used to calculate a version of the standard deviational constant, plotted in figure 16, where the price in *agent time* was sampled every 1000th *agent time* step. Just as with *trade time* the curve is smooth (cf. figure 4).

Although the *agent time* did not play an important role in the data treatment of this thesis, the results show that this time notion fulfils some basic requirements for simulating real time in financial markets.

6.8 Response function

For the simpler scenarios, such as the one in section 6.4, the response function was positive and rather constant (see figure 17), consistent with the results of Bouchaud et al. referred to in section 5.2.5. There was no sign, though, of the slight initial increase and the later decrease of the response function also mentioned buy these authors.

6.9 Standard-deviational constant and diffusivenenss

As shown in figure 4, the standard deviational constant turned out not to be constant for *ZeroIntelligenceTraders*, but heavily sloped, indicating a sub-diffusive process. The same result was given by *RandomTraders*.

Taking the reasoning of section 1.4.1 as a starting point, that is seeing diffusiveness as a balance between opposing forces, it was clear that in this setting, the sub-diffusive forces were getting the upper hand. These traders were acting both as liquidity providers and as liquidity takers, but the persistence of the order book made liquidity providing more beneficial. Since the agents did not

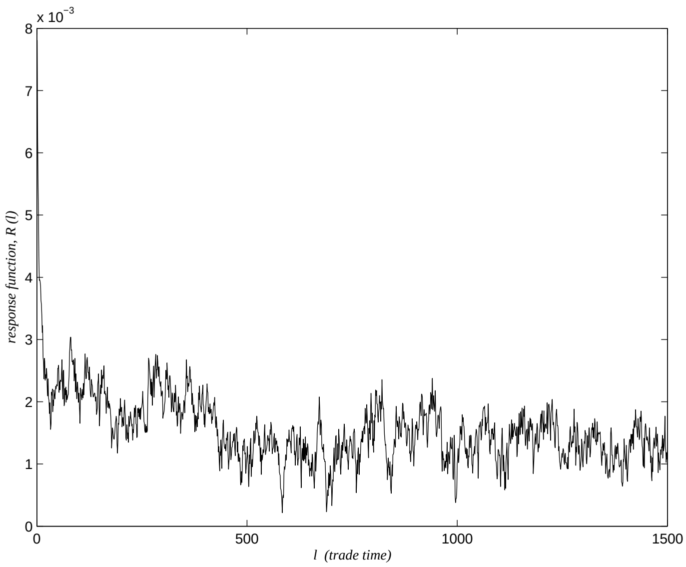

Figure 17: The **response function** from a realization over 10 000 trades with 300 *RandomTraders*, setting prices (initiated at 100.00 €) with a variance of 1.00 €.

adapt their strategies, they continued creating market orders as well as limit orders, which is unrealistic, since it actually made them lose money. In other words, the strategies of this simple kind of traders seemed to be all right for providing liquidity, but not acceptable when taking liquidity.

The solution to this problem was to use two different populations, one consisting of agents placing limit orders only, another consisting of agents placing market orders only.

The ***LiquidityProvider*** is a trader who

1. places limit orders (buy or sell with equal probabilities), where the price is taken from a normal distribution centered around the current ask if it is a sell order, around the current bid if it is a buy order,

2. refrains from creating the order if trading at market would have given a better (or equally good) deal,

3. only puts in a new order if she does not already have one of the same "sign" (buy or sell) in the order book,

4. cancels an existing order if and only if she is about to emit a new order of the opposite "sign".

Basically, this is a *RandomTrader* who chooses never to act as a liquidity taker (i.e. placing market orders).

The other agent type, the ***RandomInformedTrader***, was designed to act as an informed trader, that is with a clear conviction that either buying or selling will be a winning strategy. It is a trader who simply

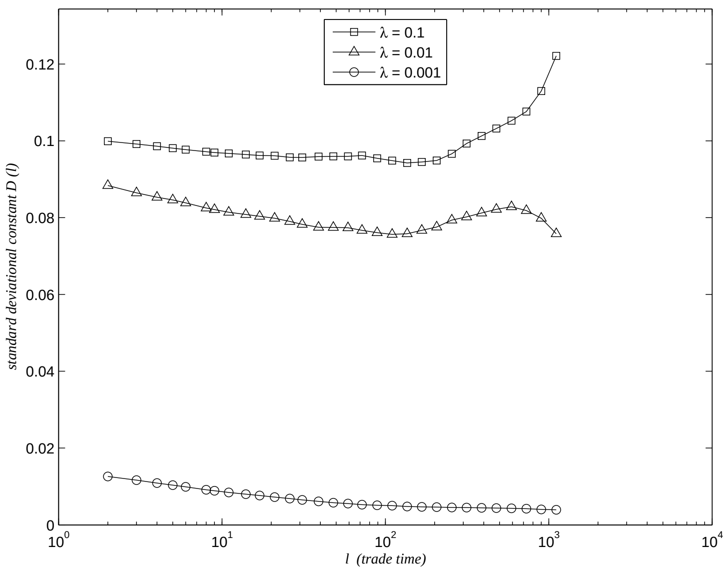

Figure 18: The **standard deviational constant** function of three different realizations, using *LiquidityProviders* ($P_{skip} = 0$) and *RandomInformedTraders*. The diffusiveness varies as the λ of the *RandomInformedTraders* assumes different values.

1. places market orders, either buy orders only or sell orders only, at random occasions.

In the realizations, 150 traders of each kind was used. However, the activity of *RandomInformedTraders* was controlled via their *likeliness to trade*, λ, a parameter used as the probability of a *RandomInformedTrader* to place a market order when in turn. A higher λ would thus make the liquidity takers more active, increasing their influence on the market at the expense of the liquidity providers.

Figure 18 shows realizations with different values of the λ parameter. When the activity of liquidity takers increases, their super-diffusive influence on the market changes the curve of the standard deviational constant, reducing the decrease and even turning it into an increase for larger time horizons.

6.10 Trade sign autocorrelation

6.10.1 Correlated series of market orders

Without elaborate strategies at work, the sign autocorrelation function should dance around zero for all l-values greater than zero. The analysis of Bouchaud et al. [7] on real trading data, however, shows significant sign autocorrelations, decaying slowly, as a power law. As a possible explanation, the authors suggest a mechanism where liquidity takers divide large market orders into several smaller ones, thereby creating the correlations needed to explain the slow decay of the sign autocorrelation.

An attempt to mimic this mechanism was made using two kinds of traders, one population of liquidity providers, that is the *LiquidityProviders* known from section 6.9, and one of a new kind of liquidity takers. The latter trader type, the ***SerialTrader***, acts solely as a liquidity taker and is a trader who

1. starts a correlated series of ten market orders (either all buy orders or all sell orders) with a probability λ_{series},
2. trades whenever she is in turn once a series is started, until the series is complete, but refuses to accept a worse deal than the first one in the series of correlated trades.

For example, a bullish *SerialTrader* would buy one share at market, note the price as her reference price and then buy another share as soon as the price was below the reference. The price constraint was added to reflect that a real trader, even if informed, probably would not trade at any price, but only at a price where the earnings would be substantial enough to justify the investment.

In order to ensure a fair game, the realizations were run with as many bulls as bears amongst the liquidity takers. The balance between *LiquidityProviders* and *SerialTraders* was controlled by adjusting the λ_{series} of the *SerialTraders*. A corresponding parameter, P_{skip} controlled the activity of *LiquidityProviders*, by making them skip their turn with this probability.

As for the results, this setup did *not* generate any sign autocorrelations, see figure 20. Although the individual liquidity takers emit series of sign correlated market orders, these series may be interlaced with those of other liquidity takers. If the market orders of different traders may be sell orders just as well as buy orders, the sign correlations will be diluted. On heavily traded stocks, such as France Télécom studied by Bouchaud et al., there is as much as one trade per second. If the liquidity takers have trading horizons spanning from, say, a few minutes to several weeks, there will be lots of trades in between the correlated trades of an individual liquidity taker, and the sign correlations will disappear. It is thus difficult to see how the mechanism of dividing up large market orders into smaller ones in itself could account for the slow decay of the sign autocorrelation.

6.10.2 Similar trading criteria

The results of the previous section imply that a slowly decaying sign autocorrelation cannot emerge, on heavily traded assets, unless several *different* traders place the same type of market orders within a short time span. If this happens, though, there are great chances that, say, a buy order will be followed by another buy order.

One way of creating a market where this condition is fulfilled, is to let many liquidity takers share *similar criteria* for determining when and how to trade. Many of them will then have the same opinion on when it is a good time to buy and when it is a good time to sell.

Two populations, similar to the ones above, were used in a crude attempt to create such a scenario. The *SerialTraders*, though, were replaced by liquidity takers who shared a common market expectation: the ***ExpectingTrader*** is a trader who

1. makes a decision whether to start a correlated series of ten market orders (either all buy orders or all sell orders) or not, with a probability dependent

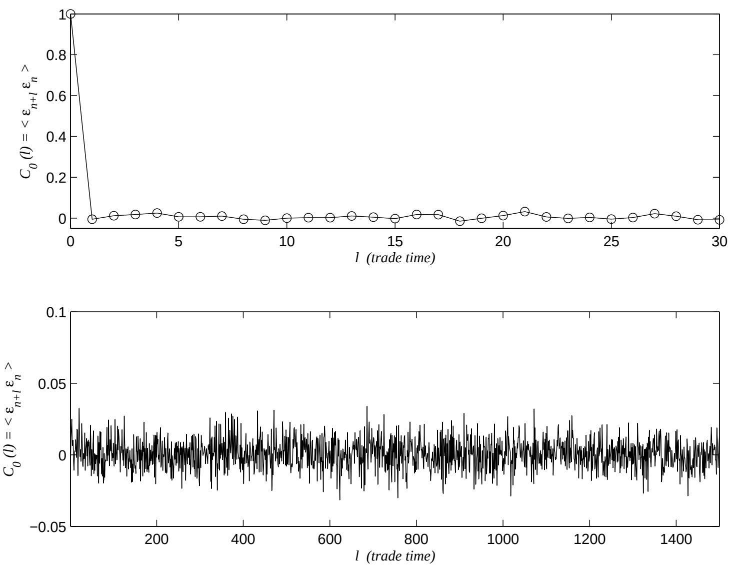

Figure 19: The **trade sign autocorrelation function** from a realization over 10 000 trades with 200 *SerialTraders* ($\lambda_{series} = 0.01$) and 100 *LiquidityProviders* ($P_{skip} = 0$), setting prices (initiated at 100.00 €) with a variance of 0.10 €. The upper diagram zooms in on l-values from 0 to 30, the lower diagram shows the development over longer time horisons, where l-values are spanning from 1 to 1500. This is clearly not a slowly decaying function.

on the distance between the current midpoint price and an expected future price (set at 102.00 € when the initial price was 100.00 €, the tick size 0.01 €, the variance of limit prices 0.10 € and the total number of trades 10 000),

2. trades whenever she is in turn once a series is started, until the series is complete, but refuses to accept a worse deal than the first one in the series of correlated trades.

Her probabilities of placing market orders was adjusted as a function of the gap between the current midpoint and the expected future midpoint, so that a large gap would be considered a bargain, while a small or a negative price gap would make the trader reluctant to place any orders. For simplicity, the fraction of the two price values was used instead of the difference, and in the case of a bullish trader and a current midpoint below the expected future price, the formula was

$$\lambda^*_{series} = \lambda_{series} \left(\frac{m}{p_{expected}} \right)^{200} \tag{19}$$

so that the likeliness to trade, λ_{series}, is enhanced by a factor depending on the fraction of the current midpoint, m and the expected future price, $p_{expected}$. This fraction was inversed either if the trader was bearish or if $m > p_{expected}$. The rather arbirary exponent 200 is of course not universal, but seemed to work fine with the parameter settings in use.

The resulting sign autocorrelation function is plotted in figures 20 and 21. Although a lot of noise is blurring the function, it is clearly a matter of slow decay, and the order of magnitude seems consistent with the analysis of Bouchaud et al.

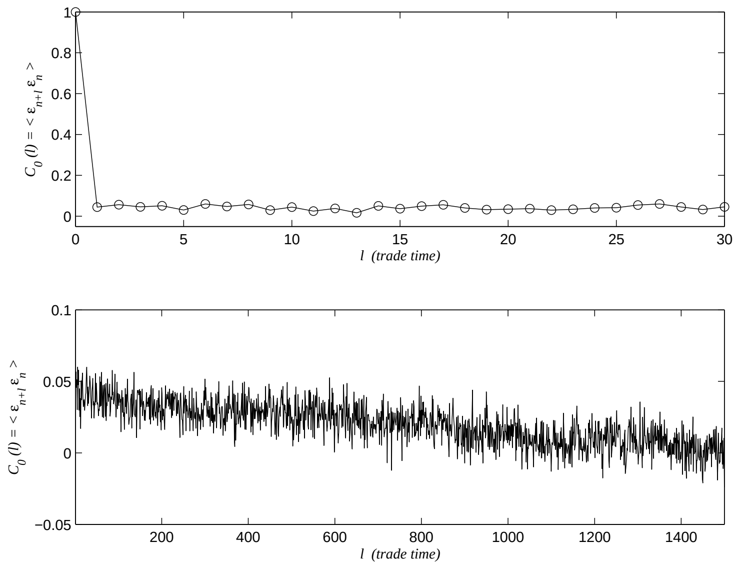

Figure 20: The **trade sign autocorrelation function** from a realization over 10 000 trades with 200 *ExpectingTraders* ($\lambda^*_{series} = 0.001$) and 100 *LiquidityProviders* ($P_{skip} = 0.1$), setting prices (initiated at 100.00 €) with a variance of 0.10 €. The upper diagram zooms in on l-values from 0 to 30, the lower diagram shows the development over longer time horisons, where l-values are spanning from 1 to 1500. Contrary to the realization of figure 19, there is evidence of correlations with a slow decay. See also figure 21
.

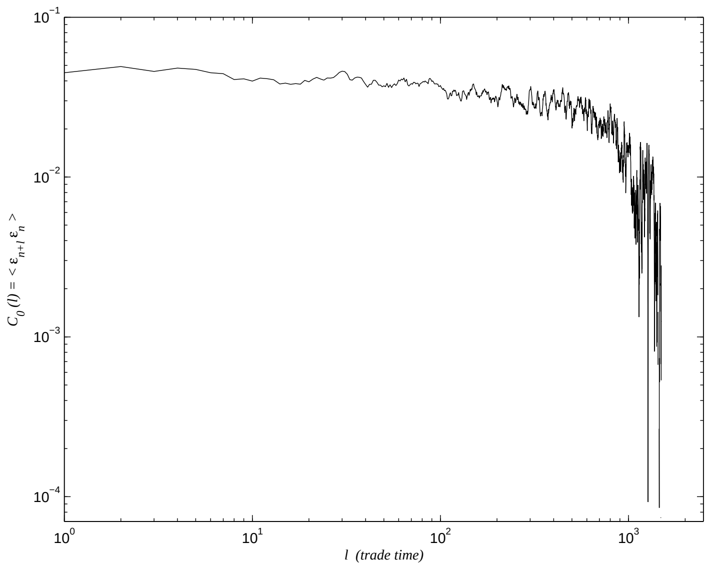

Figure 21: The **trade sign autocorrelation function** from a realization over 10 000 trades with 200 *ExpectingTraders* and 100 *LiquidityProviders*, setting prices (initiated at 100.00 €) with a variance of 0.10 €. In this log-log diagram some of the noise has been removed by smoothing the function, using an 11-point moving average.

7 Further discussion

7.1 The order book mechanism and efficient markets

The order book is at the very core of double-auction markets. It is of course interesting to see in what way this trading procedure per se influences the price development. One would perhaps expect that realizations with a poulation of zero intelligence traders would result in a standard deviational constant that actually was constant over different time horizons. However, *ZeroIntelligenceTraders* as well as *RandomTraders* rendered a declining standard deviational constant, meaning that the price develops in a sub-diffusive way.

And clearly, the order book with its storing and sorting functionalities is a mechanism that conserves the price. Smith et al. [27] states that demand storage in a non-efficient market necessarily causes *persistence*. Daniels et al. [13] also speaks of the persistence of the order book, concluding that it is non-trivial and depending on the time horizon.

When Smith et al. speaks about the persistence as a consequence of a non-efficient market, they are referring to the widespread *efficient market hypothesis* (EMH), developed primarily by Fama [15] in the 1960s. The hypothesis asserts that all actors at financial markets constantly update their offers when new information changes their perceived value of an asset. Along with certain assumptions, this would imply that the price of an asset always is “correct” with respect to all information currently available, and that it would be impossible to find assets that are mispriced, i.e. where the price differs from the “actual value" of the asset.

Assuming that markets are efficient actually predicts a market behaviour that is consistent with much of the behaviour of real markets. Clearly, though, a real market cannot be completely efficient. For example, new information will not be known instantly by all actors on a market and, as pointed out by Smith et al., offers in an order book will not be updated continuously, but will always remain there unchanged for some time.

The efficient market hypothesis has provided an important framework for economic theory, but its shortcomings call for the exploration of other approaches, such as order-book modelling.

Bovier et al. [10] makes an interesting remark that in general only the orders placed by agents are considered in order-book models. This implies that agents are visible in the model only when they place orders. However, the opinions of other agents will also affect the price development. If agents are watching the market, a price change can trigger them to place orders. So, although order-book modelling seems like a very hands-on approach that does not need to assume for example instant updating of every offer at the market, careful considerations are still needed to capture salient aspects of reality. In this particular matter of "hidden opinions", a model where agents are actually represented, such as the one used in this thesis, is preferable to order-book models where only the orders of an agent is considered, not the agent itself.

7.2 Trade sign autocorrelation

It has been shown (in section 6.10.1) that *series of correlated market orders* created by individual traders is not in itself a sufficient explanation for the empirically established slow decay of trade sign autocorrelations.

Instead a different approach was proposed, where *similar trading criteria* makes several different individuals willing to buy at the same time, or willing to sell at the same time. It was confirmed in simulation (section 6.10.2) that such a setting can indeed result in a slowly decaying sign autocorrelation function. The simple implementation used for illustrating this principle was to create groups of liquidity takers that shared a *common market expectation* that prices would raise by two percent. It is not clear whether such common expectations really exist, and the concept is closely related to that of *fundamental price*. In *fundamental analysis*, the basic idea is that the price of a share can deviate from its "correct" value in the short run, but will eventually adjust to a price that reflects the real value of the company. Bouchaud et al. make a point of suggesting that price variation is dependent not so much on any fundamental price, but rather on the mechanisms of the trading activity itself. However, similar strategies for traders could clearly arise from other reasons than a common market expectation, for example if they all make similar statistical analyses of available trading data.

7.3 Price information used for trading decisions

Traders who are thinking about buying need to decide when to buy and what the reasonable price to pay is. If they are buying at market, they will probably look at the current best ask as a reference price. If they are placing limit orders, on the other hand, they would perhaps take a glance at the best bid, to see how much the others are prepared to pay. In fact, all of the price notions defined in section 4.6.1 could be candidates for best reference.

Presumably, real traders consider price information in a more complicated manner. Prices can be *aggregated* in different ways. For example, a realization was briefly tried out where *RandomLiquidityProviders* used as a reference a weighted average of limit prices from the order book, with little consequences for the observables.

The most wide-spread form of price aggregation is of course in time: in addition to looking at the current price to pay, past prices can also be taken into consideration. Even though statistical analysis of past prices cannot predict the development of the random walk that leads future prices, charts of price history appear everywhere. The charts are studied by traders, who therefore most likely are influenced by the price history when they make their trading decisions. Unfortunately, this gut feeling inspired by the price history does not easily translate into any rational trading strategy. In this model, trading strategies were imposed and not developed endogenously. In lack of realistic (or at least rational) strategies to impose, no realizations were made with traders looking at past prices.

Also, with the basic strategies of the traders used in this thesis, the price notion used for trading decisions seemed not to be a critical issue. For example, the results were similar whether liqudity providers used the current midpoint or the current buy when placing new limit buy orders. If more elaborated strategies were implemented, though, they would probably be more sensitive to the choice of reference price.

7.4 More on the discretization of continuous time

Poisson processes are used in some applications where continuous time is modelled. In a given time window, the Poisson distribution will provide the number of events occurring as well as the positions of these events within the time interval. However, any Poisson process must fulfil the condition that the number of events occurring in two disjoint subintervals are independent random variables. This is not the case in financial markets, where a raise in activity actually increases the probability of high activity in the near future. Such *volatility clustering* becomes most apparent in the herd behaviour of investors that somtimes actually leads to a veritable crash.

Thus, a drawback with the Poisson process approach is that it will inevitabely fail to capture some of the most intriguing features of financial markets, such as avalanche effects leading to craches. On the other hand, it might give a more accurate description than this thesis of the normal, everyday trading.

7.5 Model critique and suggested improvements

7.5.1 Validating the model

The validation of the model has been centred around *descriptive output validation*. This means that the output data of the model has been compared to analysis of real data, and is considered a good model if the output data bear close resemblance to real data. Other means of model validation would probably improve the model further, especially *input validation*, where it should be verified that the structural conditions and behavioural dispositions incorporated in the model capture the salient aspects of the real system.

Also, the *robustness of the model* should be validated. Is it, like the Santa Fe stock market, sensitive to certain parameter values, and for which kinds of scenarios can the model actually be expected to reflect real market behaviour?

7.5.2 Volume

The implemetation used in this thesis has great unused potential when it comes to adding the complexity of order volume to the simulations. Volume does play an important role in real markets, which are often dominated by a few large actors, who are buying and selling large volumes at a time, affecting prices substantially. In this thesis, one share only was traded at a time, even if the implemetation could allow for more. The problem with adding volume in the scenarios lies elswere: increased complexity in the analysis of the data, but also in the trading rules of the agents. Currently, only the “sign” (buy or sell) and price of an order has to be determined, but when volume is added it is hard to determine for example what the relation should be between the price and the volume (buying small volumes at a high price could probably be more acceptable than buying large volumes at a high price, etcetera). The simplest solution for adding volume would be to let price and volume be mutually independent, and just decide the volume of each order randomly from a distribution that is consistent with empirical analyses of the order flow (e.g. [6]).

7.5.3 Order flow

The order flow, that is the sequence of orders arriving to the order book, can easily be made more realistic than in the present model. In order book models where agents are not visible, but only the incoming orders, researchers generally pay much more attention to the nature of the order flow, making sure that it is statistically similar to the order flow accounted for in empirical studies. This can be done in models using agents as well, by using more appropriate distributions for price decisions, and also by a systematic use of order cancellations. There are many empirical studies (e.g. [5]) that statistically describes the order flow, as well as the order book profile (i.e. how the volume of the limit orders are spread out across the price dimension) that can be used for comparisons.

7.5.4 Trading strategies

Since the model of this thesis uses explicit trading strategies, it can be used to try out if different proposed trading strategies, for example the detailed liquidity provider strategy suggested by Bouchaud et al. in [9], actually result in the expected market behaviour.

For liquidity takers, it would be interesting to work with fixed time horizons for investments. An informed trader has a limited period of time to buy or sell the shares concerned by the information, since the information is useless once the market has taken it into account by adjusting prices. However, is seems difficult to find rational strategies for when to invest within a given time span: if you want to buy some shares before a certain date and you notice that prices start to fall - will you then buy straight away, so that you do not risk that the price reverts to its previous value, or will you wait, hoping that this initial price decline is just the start of a more drastic price dip, which you could profit even more from by buying at a later time?

7.5.5 Agent characteristics

One of the great challenges for the agent-based modelling lies in deciding which of the agents' characteristics, constraints and abilities are worth implementing. For example, should their time to react on new information be taken into consideration? Should their information access be dependent on how much attention they are paying to the price development of the particular asset of the simulated order book? Should they be "emotional" in their decisions, for example reluctant to sell a share back at a lower price than they paid for it. Should they be realistic in the sense that they act on more specific time horizons, so that they rarely would buy a share and then sell it back a minute later?

8 Conclusion

In this thesis, a specific implementation of an artificial financial market was tried out. It was able to reproduce data that shared several characteristics of real financial data, as analysed by for example Bouchaud et al. [7].

The article of Bouchaud et al. [7] was in several respects used as a starting point for the work of this thesis, not only for the construction of the model, but also for the choice of simulation scenarios. Two of the authors' assertments were investigated more throughly:

1. **The statement that a slowly decaying trade sign autocorrelation function can be easily explained by the correlated market orders of individual liquidity takers was critizised**. A different explantion was proposed, where similar trading criteria of groups of liquidity takers could result in periods where market orders are likely to be followed by other market orders of the same "sign" (buy or sell). Although this result on trade sign autocorrelations was reached through an implementation using a common market expectation, the interesting part of it has nothing to do with market expectations: it shows that similar trading strategies (no matter what their origin may be) actually can create slowly decaying sign autocorrelations, while it is difficult to see how this pattern could appear from the correlated market orders of individual traders, as proposed by Bouchaud et al.

2. **The suggested logic that liquidity providers and liquidity takers create sub- and super-diffusive forces, respectively, that can be balanced into an overall diffusive price development, was realized in simulations**. In this implementation, where static trading strategies were imposed on the model, different kinds of traders were shown to influence prices differently with respect to diffusiveness. With the right parameter settings, the standard deviational constant became indeed nearly constant.

9 Contact

For further information, such as any related discussions or perhaps the Java source code, please contact the author (info@berseus.se). For more information on ongoing econophysics projects at the Division of Mathematical Physics of Lund Institute of Technology, please refer to Professor Thomas Guhr (thomas.guhr@matfys.lth.se).

10 Acknowledgements

I would like to thank my supervisor **Thomas Guhr** for going through many problems from a theoretical perspective and for communicating strong convictions on how scientific work should be pursued.

I would like to express my gratitude to my supervisor **Rudi Schäfer** for inspiring discussions on many issues, for helping me out with technical problems, and especially for his generous attitude of never imposing his own solutions and ideas, but instead allowing me to develop my own lines of thought.

Thanks to my examinator for taking the time to read this report and to attend the oral presentation of my thesis.

Thanks to everyone at the Division of Mathematical Physics for making my time there a pleasant one.

References

[1] R. Axelrod & L. Tesfatison (2005). *A Guide for Newcomers to Agent-based Modeling in the Social Sciences.* Appendix in Handbook of Computational Economics, Vol. 2: Agent-based Computational Economics.

[2] L. Bachelier (1900). *Théorie de la spéculation*, Annales scientifiques de l'.N.S. 3e srie, tome 17, 21-86.

[3] B. Biais, P. Hillion & C. Spatt (1995). *An Empirical Analysis of the Limit Order Book and the Order Flow in the Paris Bourse.* The Journal of Finance 50 (5), 1655-1689.

[4] F. Black & M. Scholes (1973). *The Pricing of Options and Corporate Liabilities.* Journal of Political Economy 81 (3), 637-654.

[5] J.P. Bouchaud, M. Mézard, M. Potters, Statistical properties of stock order books:empirical results and models, Quantitative finance 2, 251 (2002).

[6] J. P. Bouchaud, J. Mézard & M. Potters (2002). *Statistical properties of stock order books: empirical results and models.* Quantitative Finance 2, 251-256.

[7] J.-P. Bouchaud, Y. Gefen, M. Potters & M. Wyart (2004). *Fluctuations and response in financial markets: the subtle nature of 'random' price changes.* Quantitative Finance 1, 223.

[8] J.-P. Bouchaud (2005). *The subtle nature of financial random walks*, Chaos 15, article 026104.

[9] J. P. Bouchaud, J. Kockelkoren & M. Potters (2006). *Random walks, liquidity molasses and critical response in financial markets.* Quantitative finance, 6, 115.

[10] A. Bovier, J. Černý, O. Hryniv (2006). *The Opinion Game: Stock price evolution from microscopic market modelling.* International Journal of Theoretical and Applied Finance 9 (1), 91-112.

[11] P. Callahan (accessed on February 5, 2007). *What is the Game of Life?* URL: http://www.math.com/students/wonders/life/life.html.

[12] N.T. Chan, B. LeBaron, A.W. Lo & T. Poggio (1999). *Agent-based models of financial markets: A comparison with experimental markets.* Working Paper for MIT Artificial Markets Project.

[13] M.G. Daniels, J.D. Farmer, G. Iori & E. Smith (2002). *How storing supply and demand affects price diffusion.* Technical report, Santa Fe Institute.

[14] A. Einstein (1905). *Uber die von der molekularkinetischen Theorie der Wrme gefordete Bewegung von in ruhenden Flssigkeiten suspendierten Teilchen.* Annalen der Physik 17, 549.

[15] E. F. Fama (1970). *Efficient Capital Markets: A Review of Theory and Empirical Work.* The Journal of Finance, 25 (2), 383-417.

[16] D.K. Gode and S. Sunder (1993). *Allocative Efficiency of Markets with Zero-Intelligence Traders: Market as a Partial Substitute for Individual Rationality.* Journal of Political Economy 101, 119-137.

[17] T. Guhr (2005). *Econophysics.* Matematisk Fysik, LTH, Lunds Universitet.

[18] P.E. Johnson (2002). *What I learned from the Artificial Stock Market.* Social science computer review 20 (2), 174-186.

[19] B. LeBaron (2000). *Agent-Based Computational Finance: Suggested Readings and Early Research.* Journal of Economic Dynamics and Control 24 (5-7), 679-702.

[20] B. LeBaron (2002). *Building the Santa Fe Artificial Stock Market.* Working Paper, Brandeis University.

[21] B. LeBaron (2005). *Agent-based Computational Finance.* Working Paper, Brandeis University.

[22] F. Lillo & J.D. Farmer (2004). *The long memory of efficient market.* Studies in Nonlinear Dynamics and Econometrics 8, 1.

[23] L. Tesfatsion (2002) *Agent-Based Computational Economics: Growing Economics from the Bottom Up.* ISU Economic Working Paper No.1. Iowa State University.

[24] L. Tesfatsion (2003) *Agent-Based Computational Economics.* ISU Economic Working Paper No.1. Revised August 24, 2003. Iowa State University.

[25] L. Tesfatsion (2006), *Agent-Based Computational Economics: A Constructive Approach to Economic Theory*, Handbook of Computational Economics, Volume 2: Agent-Based Computational Economics.

[26] M. Potters & J.-P. Bouchaud (2003). *More statistical properties of order books ans price impact.* Physica A: Statistical Mechanics and its Applications, 324 (1-2), 133-140.

[27] E. Smith, J.D. Farmer, L. Gillemot & S. Krishnamurthy (2003). *Statistical theory of the continuous double auction.* Quantitative Finance 3, 481-514.

[28] Wikipedia (online encyclopedia, accessed on February 5, 2007) *Continuous-time Markov process.*
URL: http://en.wikipedia.org/wiki/Continuous-time_Markov_process.

[29] Wikipedia (online encyclopedia, accessed on February 5, 2007) *Moment (mathematics).*
URL: http://en.wikipedia.org/wiki/Moment_%28mathematics%29.

[30] M. Wyart, J.-P. Bouchaud, J. Kockelkoren, M. Potters & & M. Vettorazzo (2006). *Relation between Bid-Ask Spread, Impact and Volatility in Double Auction Markets.* Science & Finance (CFM) working paper archive 500067.

A Appendix: Glossary

agent-based modelling	A bottom-up approach where a large population of simple agents interact on a microscopic level, so that state variables emerge on a macroscopic level.
artificial market	A market situation that is created artificially, either by giving a group of human beings specific instructions and then let them trade with each other, or by implementing a computer program that simulates a market, with more or less intelligent agents or order flows.
ask price	The price specified in a limit sell order.
asset	Something of value. Can either be a commodiy or a security.
bear	A trader who is trading on the expecatition that prices will drop. A bear is the opposite of a bull.
bid price	The price specified in a limit buy order.
bull	A trader who is trading on the expecatition that prices will raise. A bull is the opposite of a bear.
bid-ask spread	The price gap between the best bid and the best ask in the order book.
diffusion	A motion where objects are scattered in a random-walk manner, or, to put it in technical terms, that the second moment with time is linear.
diffusion constant	The proportionality constant of the linear relation between the volatility and the elapsed time in a diffusive process.
double-auction market	A market where both buyers and sellers have the opportunity to compete with their offers, i.e. there will be, at the same time, both buy offers and sell offers, which are often handled by a central order book.
efficient market hypothesis	A theory that asserts that all known information is included in the price of an asset, implying that it is impossible for any participant to consistently outperform the market.

emergence	The idea that simple behaviour on a microscopic level of a system can result in complex or unexpected behaviour on a macroscopic level. In this thesis, simple strategies on the trading level of a financial market are introduced, while the statistical analysis is made on the price curve at market level. Although the concept of emergence itself has something of a quasi-scientific reputation, it illustrates how physicists can offer useful insights to the world of finance: being used to study complex systems at both microscopic and macroscopic levels, physicists can use methods and models originally developed within statistical physics to analyse the processes occurring in financial markets.
financial market	A mechanism that allows traders to trade money for securities or commodities. A stock exchange is an example of a financial market.
fundamental analysis	A way of analysing market data that builds upon the assumption that there exists a correct price for each asset, and that the market in the short run may misprice the asset, but that it will eventually correct for its mistake by elimination of the misprice.
hedging	Actions to reduce the risk in financial investments. An example is when you invest in several companies of fundamentally different lines of business instead of putting all your eggs in one basket.
informed trader	A trader that tries to make a profit on (what the trader believes is) superior information. Will generally act as a liquidity taker.
limit order	A buy or sell order that specifies a limit price, so that it cannot be matched with other orders offering a worse deal.
liquidity	A measure of how easily matching orders appear at a market. A narrow bid-ask spread and large volumes available in the order book are signs of high liquidity. If the liquidity is too low, the trading activity comes to a halt, which is why market organizers are very concerned with keeping the market as liquid as possible.
liquidity taker	A trader, possibly informed, that has decided to buy or sell securities. In order not to inflict suspicison, liquidity takers tend to divide large orders into several small ones.

liquidity provider	A trader placing limit orders and making profit not on future price expectations, but rather on the difference between the best buy and the best ask.
market maker	An actor on a market that undertakes to place buy and sell orders, making sure that the bid-ask spread is not too large for a certain security, thereby creating liquidity in the market for this security.
market order	A buy or sell order that accepts the conditions of the best offer amongst the current limit orders, and therefore is executed immediately.
Markov process	A stochastic process with the Markov property, meaning that future states depend only on the current state of the process, and not on past states.
order book	The current set of active limit orders in a market, grouped into buy orders and sell orders and generally sorted according to price-time priority, i.e. first according to price and then, if needed, by the time of appearance in the order book.
Poisson process	A statistical model used to determine how closely random events occur in time or in space. Often used for continuous time simulations.
quote	A record that briefly shows the state of the order book: best bid, best ask, available volume and a time stamp.
random walk	The path that arises if successive steps are taken, each in a random direction. The discrete version of Brownian motion.
security	An asset in the form of a formal declaration, which gives the holder the right to receive interest or dividends. Common examples are bonds, equities (shares) and derivatives.
stochastic process	A sequence of function values that depend on probability distributions. Financial time series are often modelled as stochastic processes.
tick size	The minimal price difference on a market. The price of an order can only be as precise as the tick size.

trade An exchange of assets for money, which is described in the data by traded price, traded quantity, and a time stamp.

trade time A time notion where each trade makes the time progress by one step, regardless of the actual time that passed since the last trade.

trading time The time when a market is open and trading is possible.

volatility A quantification of how widely spread the changes in value of a financial instrument are. In statistical terms, the volatility is often defined as the standard deviation of the change in value within a specific time horizon.