POLITECNICO DI TORINO

Double MASTER's Degree in ***Physics of Complex Systems***

Master Thesis

Herding Behaviour in Latent Order Books

Supervisor

Dr. Luca DALL'ASTA

Candidate

Giovanni Michele CHERCHI

Co-Supervisor

Dr. Michael BENZAQUEN

Academic Year 2019-2020

Abstract

Market micro-structure is a branch of finance describing the details of how transactions happen in the order book between buyers and sellers and what is their effect on price formation. Furthermore, the fundamental laws of supply and demand are not avoidable, meaning trading always has a cost or an impact on the price. Economic equilibrium happens when the quantities demanded, at the current price, are equal to the quantities supplied, and this holds also for liquid assets in financial markets. Investors are interested having the best strategies and planning the investments (*meta-orders*) and this motivated extensive research in the academic world, addressing the problem of building a valid mathematical model to reproduce empirical impact curves emerging from the underlying micro-structure. Thus, this implies that endogenous effects play a vital role, justifying also the *Universality* associated to the mentioned empirical laws.

Starting from existent mathematical models, the present work aims first to implement a robust computational framework able to reproduce the dynamics of a Latent Order Book (LOB) and its transaction cost curves. Secondly, it tries to address the problem of herding behavior among traders and how this reflects on the impact cost, called *co-impact*, and price correlation, comparing with some additional recent empirical results.

Acknowledgements

I would like to thank my parents for the support, for having believed in me despite all the difficulties. Without you, any of my accomplishments would have been possible.

I would also like to thank my friends in Turin, with whom I spent joyful moments throughout this journey.

Special thanks to Eugenio, who helped me sort out many crucial problems. Not everybody would have done that and therefore, thank you!

To Tania, for being there when I needed, for encouraging me to overcome many obstacles, for the good advices, for the laughs, thanks for putting up with me!

Table of Contents

Chapter 1

Introduction

1.1 Overview

In market micro-structure, the study of how the price of financial assets [1] are affected by planned execution of a certain volume of orders Q, called *meta-orders*, has become increasingly important throughout these last years.
In fact, there are very well established empirical results (see e.g. [**1**], [**2**]) that proved the existence of *Universal laws* regulating the mechanics of an order book[2].
This aspect has an important implication that, perhaps, goes against a common misconception: financial markets are not as susceptible by global news media as one may think.
Indeed, most of price movements are exclusively an endogenous effect which derives from the market micro-structure rather than events in the world's political or economic environment, for instance, see [**3**], [**4**].
As in many problems in theoretical physics, *Universality* allows one to abstract from microscopic details, and focus on global parameters that describe the system; this means constructing a *"coarse-grained"* model that should be able to reproduce the above mentioned law. Therefore, in the price dynamics there should be some large scale phenomena emerging from the order book's microscopic dynamics.
Furthermore, *Universal Laws* regarding the price imply that its behavior is independent on the period in which the exchange happened, the geographical position or even the kind of contract traded (Bitcoin, options, futures, ...).

[1]Any resource owned by a company or multinational. It can produce positive economical value and can be converted to cash(e.g. Apple, Google, etc,..).

[2]List of orders that records the interest of Buyers and Sellers of a particular asset or financial instrument. A matching tool is used to determine which orders can be fulfilled and at which price.

The total orders visible, daily, are much less than total Market Capitalization [3], and the quantity which is visible in the order book is again many orders of mangnitude smaller.

Subsequently, the *latent order book* has been introduced, which contains the intentions of the traders in the market that are not yet visible in the *revealed book*, playing an important role in the formation of the price.
Mathematical models have been built to explain this phenomena [**5**], together with more intense empirical research in order to understand the extent of validity of the square-root law.
More recently, there is evidence of a square-root/linear cross-over behavior of price returns and this is well described in the models of latent order books [**6**], but it would be interesting to see this in our simulation, even if unfortunately, it does not seem the case.

1.2 Aim of the Research

There are still some open problems for what concerns the average price impact in presence of multiple meta-orders execution; traders may in fact have different strategies that result in different execution frequencies.
Plus, meta-orders are in general correlated both in size and sign (*buy* or *sell*) in different ways.
The first goal of the research is to built a solid computational framework, which is able to reproduce the principal features of the latent order book in the linear approximation. Furthermore, it tries to include for the effect of noise traders, investigating its effects on the impact curve and relaxation.
The simulation accounts also for called *crowding effects* or *herding*: many different traders plan orders' execution following another trader for which the strategy is known; consequently, the other traders follow, imitating the same behavior.
Firstly, we simulated the latent book in presence of only one meta-order type with a given frequency of execution: this reproduces the square-root law as predicted by the model, which is a reaction-diffusion process physics' inspired.
In second place, we substituted the lonely trader with the multiples correlated agents, reproduced by noise traders with a given correlation exponent suggested by the empirical evidence.
Third, we tried to simulate the co-impact of noise traders plus a known constant

[3] It represents the total market value of an asset (maybe owned by a third company) and it is obtained multiplying the current price of one share or order, times the total number of shares.It can be expressed in different currencies.

frequency agent.

1.3 Methods

For obvious reasons, the implementation of a continuous mathematical model is a discrete process in which each orders is independent of the other and diffuses in the price-axis following the rules given by the master equation. The price axis was initialized with a certain shape(*linear*) which ensures the stationarity of the process.
The simulation was made using the scripting language *Python*.
The code was also parallelized in order to optimize the execution time and there has been an extensive use of servers provided by Ecole Polytechnique at Ladhyx.
Each plot was made making many realization of the same process and then averaging over this realizations, recording the *mid-price* time series.

1.4 Plan of the thesis

The work is divided in three parts. First, an introduction to the order book concept, secondly the mathematical model of the latent order book and the further modification to account herding behaviours and co-impact behavior.
Thereafter, we show the results of the simulation, from the already known model to the modified one. Plus, we show some results of the co-impact behavior, compared with empirical results.

Chapter 2

The Order Book

2.1 What is an order book?

The order book is the principal tool used in financial markets to satisfy *supply* and *demand*.
They are used in every *stock exchange* around the world.
In the old days it used to a paper list, now it is far more complicated that that, but the principle has never changed: for a given *price level*(or price tick), it records the interest of investors on a particular *security* or asset.
The list is divided into two: the `bid`(*demand*) side and the `ask`(*supply*) side for every price.
In figure 2.1, a sketch of an order book is depicted.

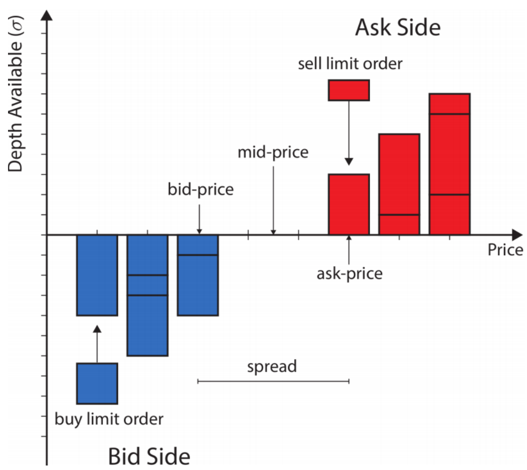

Figure 2.1: Sketch of a generic order book.

The ask and bid are well represented and also the *mid-price*, halfway from the best buy price and the best sell price.
Besides the *bid* and *ask* label, the order are labeled also as *limit orders* and *market orders*.

Limit Orders are not meant to be executed right after placing. Therefore the price level can be changed dynamically, given a minimum and a maximum price at which one desires to buy/sell. They may be placed at best buy/sell price or not and they can get executed at any price given that the investor wants to sell. Orders are not executed directly from investors, but they are managed by a a third, the market maker, which mediates between sellers and buyers and makes sure the market is always liquid, with a low spread.
Market Orders are executed right at the current price. meta-orders, which we will talk about next, are market orders.
The investor shows interest in a particular stock and the broker instantly sells or buys orders. Of course there may be some delays, due to available liquidity or anticipated stock exchange closing time.
Fundamental rule of supply and demands wants that orders are fulfilled at the price which intersect buy orders and sell orders: providing more liquidity on one side makes the price push in the opposite direction. The change in price is called impact and works in the same way both for *market* and *limit* orders.
The surprising fact is that for *meta-orders*, the *impact* follows a **Universal** law, called the *square-root* law. In physics, whenever there is some universality, this comes from some underlying low-frequency actors. These actors are the traders who form the order book.

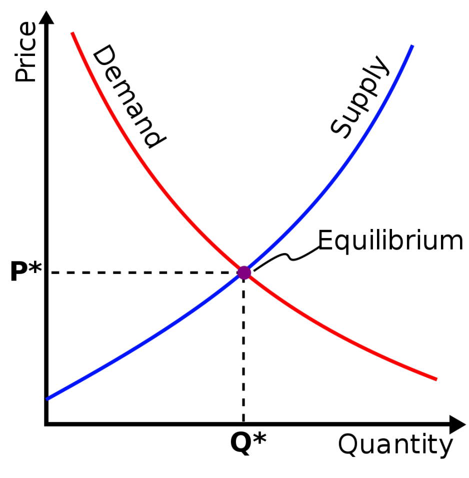

Figure 2.2: Supply and Demand Curve. P^* is the transaction price or *equilibrium price* and Q^* is the equilibrium quantity traded.

2.2 Revealed order book vs. Latent order book

We have argued that the order book records the traders' placing orders in a form of a list. For a given asset, we know that the number of orders actually traded daily is $\frac{1}{200}$ smaller than the market capitalization.
And the number of orders actually visible are 1000 less than that.

Therefore, the role of the *market maker*, as mentioned before, is provide continuous liquidity to the *revealed order book*, which is nevertheless, much lower that the liquidity actually recorded by **intermediaries** of any kind.
In addition, the orders are revealed only close to the current *transaction price/mid-price.* This large liquidity somehow hidden can be resolved only in longer time-scales than characteristic time-scales of the revealed order book.
At the same time, *low frequency/long time-scales* actors justify the *"coarse-grained"* model description and the presence of *Universal laws* dominating the dynamics of the price.

Thus, that is how the idea of ***latent order book*** emerged. It should contain the intentions of the traders, recorded somewhere, but not directly visible. This fact has been a weak spot for critics but its inference has proved to be feasible recently (see [**11**]).

The dynamics of the *latent order book(LOB)* can be modeled as a reaction-diffusion process and should reproduce the formation of the price and the shape of the LOB.
Indeed, We can write the *Fokker-Planck equation* and look for stationary solution that can be simulated, starting from the micro-structure and collecting average signals.

Chapter 3

The Mathematical Model

3.1 The Latent Order Book

The starting point for this research is the reaction-diffusion process used to model the latent order book(LOB) shown in [**5**].
As previously said, the orders diffuse in the price axis and when a buy order matches a sell order they get executed at the that price just like a reaction process, so one has *A+B -> 0*.
As shown below in 3.1, each order is labeled as `buy` or `sell` order and they are all limit orders, which means that they do not get executed right after placing, but the traders have time to change their mind and they can in this way diffuse, raising or lowering the price, with diffusion coefficient D.
The orders can also be cancelled with rate ν and deposed with rate λ

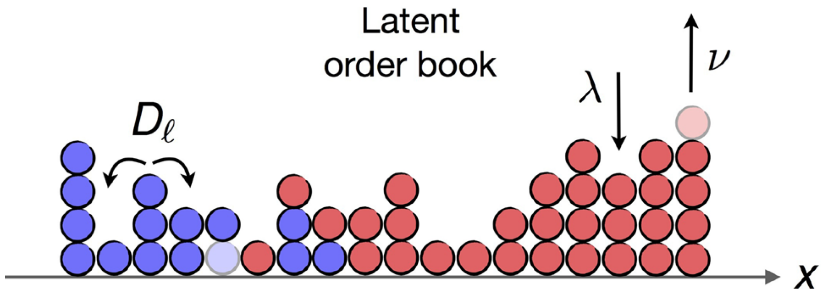

Figure 3.1: Latent book: buy (blue beads) and sell(red beads) orders are labeled. Also deposition and cancellation rates are shown as well as the diffusion coefficient.

If the size of the system goes to infinity (hydrodynamic limit), one can use the average density of orders for the bid(buy) and ask(sell), $\rho_B(x,t)$ (resp. $\rho_A(x,t)$) and write the Fokker-Planck Equations that govern the dynamic of the system:

$$\frac{\partial \rho_B(x,t)}{\partial t} = -V_t \frac{\partial \rho_B(x,t)}{\partial t} + D \frac{\partial \rho_B(x,t)^2}{\partial t^2} - \nu \rho_B(x,t) + \lambda sgn(p_t - x) - \kappa R_{AB}(x,t)$$

$$\frac{\partial \rho_A(x,t)}{\partial t} = -V_t \frac{\partial \rho_A(x,t)}{\partial t} + D \frac{\partial \rho_A(x,t)^2}{\partial t^2} - \nu \rho_A(x,t) + \lambda sgn(p_t - x) - \kappa R_{AB}(x,t)$$

The drift-diffusion term contains a shift term V_t of the entire book that accounts for some new information that is visible to all traders in the market, affecting them collectively. The prices are diffusive in reality, so this term can also be a Gaussian noise.
We can also notice that the reaction term provides a non linearity with reaction rate κ.
In fact, $R_{AB} \approx \rho_B(x,t)\rho_A(x,t) + fluctuations$.
One also takes the limit in which $\kappa \to \infty$, so the orders disappear immediately from the book when they react.
In order to get rid of the non-linearities in the equations we notice that we can define $\phi(x,t) = \rho_B(x,t) - \rho_A(x,t)$.
This quantity vanished when $x = p_t$ which is the *average reaction price* and also solution of the following PDE; this is in general different from the instantaneous one $p_t^{inst} = (a_{best} + b_{best})/2$ but we neglect this difference because is relatively small and is called diffusion width.
To sum up, we can write the new equation:

$$\frac{\partial \phi(x,t)}{\partial t} = -V_t \frac{\partial \phi(x,t)}{\partial t} + D \frac{\partial \phi(x,t)^2}{\partial t^2} - \nu \phi(x,t) + \lambda sgn(p_t - x)$$

Integrating over the informational drift component, introducing a new price $\hat{p}_t = \int_0^t V_s ds$ and changing reference frame to $y = x - \hat{p}_t$, we obtain

$$\frac{\partial \phi(y,t)}{\partial t} = D \frac{\partial \phi(y,t)^2}{\partial t^2} - \nu \phi(y,t) + \lambda sgn(p_t - \hat{p}_t - y)$$

Starting from symmetric initial conditions such that $\phi(y{,}0) = -\phi(-y{,}0)$, $\quad \hat{p}_t = p_t = 0$ is always a solution since it has to be stationary at $t \to \infty$ such that $\frac{\partial \phi(y,t)}{\partial t} = 0$ and therefore has to keep the symmetry from the origin.
In addition, we place ourselves in infinity memory limit when no cancellation and deposition of orders occurs: there are only *low-frequency* slow agents, that do not contribute immediately to the revealed liquidity in the reveled book, like

high-frequency traders do.
In this sense, the latent book serves as a *reservoir* for the revealed book.
The solution of the differential equation above is:

$$\phi_s(y \leq 0) = \frac{\lambda}{\nu}(1 - e^{\gamma y}) \quad and \quad \phi_s(y \geq 0) = -\phi_s(-y \leq 0) \quad with \quad \gamma = \sqrt{\frac{\nu}{D}}$$

This expression represents the average shape of the book and it can be further simplified in the infinite memory limit, that is, deposition and cancellation rates tend to zero.

3.1.1 Linear Latent Order Book

Near the transaction price, the approximation leads to a linear shape of the book with density given by $\phi_s(y) = -\frac{J}{D}y$, taking $\gamma \to 0$. That is the LLOB. J is the current and it is defined as $J = D\partial_y\phi_s \mid = \frac{\lambda}{\nu}$ It can be introduce also another quantity, formally the slope of the linear average shape of the book, which is the latent liquidity:
$\mathcal{L} = \frac{J}{D}$ that provides a measure of how big is the deposition rate λ with respect to the cancellation rate ν, proportional also to the diffusion coefficient: higher the deposition rate, higher the liquidity provided to the latent book, etc.

3.2 Meta-orders

We now introduce the meta-orders in the model's description. We remain in the linear approximation of the latent Order Book: if the resulting impact shifts the price by a little amount, that is justified.
That is true if the meta-order execution lasts for hours, although, for longer executions, we have to take into account nonlinearities and fewer approximations are necessary.
The meta-orders are modeled as an extra "current" of orders which arrive at the transaction price p_t, and immediately provide liquidity in the revealed book.
At the transaction price we have that y becomes $y_t = p_t - \hat{p}_t$ and so we can write the simplified differential equation for the LLOB, with an extra term that accounts for the meta-order.

$$\frac{\partial\phi(y,t)}{\partial t} = D\frac{\partial\phi(y,t)^2}{\partial t} + m_t\delta(y - y_t) \quad and \quad \frac{\partial\phi(y \to \ \pm\infty, t)}{\partial t} = -\mathcal{L}$$

The condition to impose is, of course, that the shape of the book is linear away from the price and equal to the latent liquidity $\mathcal{L}$. The solution of this equation is the same as the one previously obtained for the LLOB, plus a convolution term that depends on the form of m_t.

$$\phi(y,t) = -\mathcal{L}y + \int_0^t ds \frac{m_s}{\sqrt{4\pi(t-s)}} e^{-\frac{(y-y_s)^2}{4D(t-s)}}$$

At the transaction price y_t, $\phi(y_t, t) = 0$. One obtains therefore a stochastic integral equation for the price:

$$y_t = \frac{1}{\mathcal{L}} \int_0^t ds \frac{m_s}{\sqrt{4\pi(t-s)}} e^{-\frac{(y_t-y_s)^2}{4D(t-s)}}$$

If the impact is small such that $|y_t - y_s|^2 << D(t-s)$, one can neglect non-linearities and the mid-price becomes simply the result of a *linear propagator model*.

$$y_t = \frac{1}{\mathcal{L}} \int_0^t ds \frac{m_s}{\sqrt{4\pi(t-s)}}$$

This approximation is therefore valid for small trading rates that do not push the price that much.

3.2.1 Square Root Impact

The solution of the stochastic equation y_t provides already an *average* price signal, result of the *coarse-graining* process that the model does.
If one trades meta-orders the price will be shifted up($m_s > 0$) for buy meta-orders and down($m_s < 0$) for sell ones. Consequently, the impact depends on the quantity traded Q and can be therefore defined as:

$$\mathcal{I}(Q) = \mathbb{E}[\epsilon(y_{t+T} - y_t)|Q]$$

where T is the execution time and ϵ is sign of the meta-orders (`buy` or `sell`).
The simplest case one can imagine is constant rate of meta-orders $m_0 = Q/T$. In that case the equation for the price becomes an easily soluble integral:

$$\mathcal{I}(Q) = \frac{m_0}{\mathcal{L}\sqrt{4\pi}}\sqrt{t}$$

This equation will be useful when we will try to compare the simulation with the mathematical model. Also, it is important to underline that this holds for very low participation rates m_0.
To be more precise the integral equation for y_t has been solved, in the paper by

Donier et al. [**5**], but that is more involved and it will be reported here only the result for low and high constant participation ratio.

$$\mathcal{I}(Q) \approx \frac{m_0}{J\pi}\sqrt{\frac{Q}{\mathcal{L}}} \qquad for \qquad m_0 << J$$

$$\mathcal{I}(Q) \approx \sqrt{\frac{2Q}{\mathcal{L}}} \qquad for \qquad m_0 >> J$$

It is remarkable that for high participation ratios the Impact does not depend anymore on the execution time T, but only on volume traded, independently of the execution rate.
It is often written as $\mathcal{I}(Q) = Y\sigma\sqrt{\frac{Q}{V}}$, where Y is an order one parameter, σ is the daily volatility, which is often interpreted also as the standard deviation of the time series of the price returns during one trading day.
$V = JT_d$ is the daily traded volume.

3.3 Correlated meta-orders: herding

So far we have discussed simple and already known cases for which the *square-root* law has been extensively proved empirically.
One can ask if the law holds when correlated meta-orders are superposed to the constant rate volume.
In this case, if that constant rate trader wanted to estimate his own impact, the only known variable would be its own execution rate.
The other traders modeled with δm_s, random variable, try to copy his behaviour and trade around m_0, with a characteristic *time-delay*:

$$m_s = m_0 + \delta m_s, \qquad \mathbb{E}[\delta m_s \delta m_{s'}] = \sigma^2 (s - s')^{-\mu} \qquad and \qquad \mathbb{E}[\delta m_s] = 0$$

Since we look at increments of fractional Brownian motion, called fractional Gaussian noise, as shown in [8], the exponent μ is related to the *Hurst exponent* characteristic of Brownian motion through the relation

$$\mu = 2(1 - H)$$

The general equation for the price now becomes

$$y_t = \frac{1}{\mathcal{L}}\int_0^t ds \frac{m_0}{\sqrt{4\pi(t-s)}} e^{-\frac{(y_t - y_s)^2}{4D(t-s)}} + \frac{1}{\mathcal{L}}\int_0^t ds \frac{\delta m_s}{\sqrt{4\pi(t-s)}} e^{-\frac{(y_t - y_s)^2}{4D(t-s)}}$$

If the impact is small, in the *linear propagator model* regime we can write

$$y_t = \frac{1}{\mathcal{L}} \int_0^t ds \frac{m_0}{\sqrt{4\pi(t-s)}} + \frac{1}{\mathcal{L}} \int_0^t ds \frac{\delta m_s}{\sqrt{4\pi(t-s)}}$$

Now applying the expected value with respect to the noise, given that $\mathbb{E}[\delta m_s] = 0$, one has

$$\mathbb{E}[y_t] = \frac{1}{\mathcal{L}} \int_0^t ds \frac{m_0}{\sqrt{4\pi(t-s)}}$$

One comes back to the previous results, which implies that the noise does not affect the behavior of the impact at least for what concerns the first moment.
We will see if this approximation holds also in the simulation

Trivially, if $m_0 = 0$ then $\mathbb{E}[y_t] = 0$

The price auto-correlation reads instead

$$\mathbb{E}[y_t y_{t'}] = \frac{\sigma^2}{4\pi L^2} \int_0^t \int_0^{t'} ds ds' \frac{(s-s')^{-\mu}}{\sqrt{(t-s)(t'-s')}} e^{-\frac{(y_t - y_s)^2}{4D(t-s)} - \frac{(y_{t'} - y_{s'})^2}{4D(t'-s')}}$$

Again, in the linear propagator framework and putting $t = t'$ to obtain the variance, one has

$$\mathbb{E}[y_t^2] = \frac{\sigma^2}{4\pi \mathcal{L}^2} \int_0^t \int_0^t ds ds' \frac{(s-s')^{-\mu}}{\sqrt{(t-s)(t-s')}}$$

The variance if a function of t and we want to extract this information in order to check both the correctness of the simulation and the relation between the correlation and variance exponent. We can do a change of variable in order to have integrals from 0 to 1: setting the scaling, $s = tx$ and $s' = ty$, yields:

$$\mathbb{E}[y_t^2] = t^{1-\mu} \frac{\sigma^2}{4\pi \mathcal{L}^2} \int_0^1 \int_0^1 dx dy \frac{(x-y)^{-\mu}}{\sqrt{(1-x)(1-y)}}$$

The only problem here is the integral, that despite being a constant, possesses a divergence for x and y close to 1.
Anyway we can call this proportionality factor $\mathcal{C}$.

$$\mathbb{E}[y_t^2] = \mathcal{C} \frac{\sigma^2}{4\pi \mathcal{L}^2} t^{1-\mu}$$

Chapter 4

Simulation

This part of the report is dedicated to the simulation of the reaction-diffusion process reproducing the latent order book with the correlated meta-orders.
It will be subdivided in four parts:

- The simulation of the Linear Latent Order Book(LLOB) with constant m_0
- *LLOB* with $m_0 = 0$ and noise only meta-orders
- The simulation of the *herding behaviour*: constant rate + noise
- The simulation of the co-impact plot with respect to the participation ratio of the number of orders traded, ϕ

4.1 LLOB Simulation

The order book is initialized symmetrically (as shown in 3.1) with a linear shape in the number of orders(*limit orders*), starting from the reference initial *mid-price* $p_0 = 0$ which is defined, at any time t, as

$$p_t = \frac{1}{2}(b_t + a_t)$$

b_t is defined as the maximum price at which *investors* are willing to buy orders, while a_{best} is the minimum price at which *investors* are willing to sell them.
The *spread* is given by $a_t - b_t$.
At each step, orders independently diffuse in the price axis with the same probability to be placed at an higher or lower price, due to the fact that each *investor* has the same amount of knowledge about the future *mid-price*.
This ensures the average mid-price stays constant with respect to the initial condition, in absence of external perturbations.

The number of orders that have to move at each step and each price tick is a random variable, distributed as a binomial distribution(scheme in 3.1).
After that, each order moves *up* or *down* along the price axis, independently and symmetrically with probability, $p_{move} = \frac{1}{2}$.
Given a generic time-step t and tick size i belonging to the price axis and calling the number of selected orders $N^i_{move,t}$, the number of total orders N^i_t and the number of orders that raise or lower the price of one price tick, $N^i_{up,t}$ and $N^i_{down,t}$ respectively, we can write the following relations

$$N^i_{move,t} \sim \mathtt{Bin}(N^i_{t-\delta t}, p) \quad with \quad p = 2D\delta t$$

$$N^i_{up,t} \sim N^i_{down,t} \sim \mathtt{Bin}(N^i_{move,t}, \frac{1}{2})$$

Therefore, $N^i_{move,t} = N^i_{up,t} + N^i_{down,t} \quad \Rightarrow \quad < N^i_{up,t} >=< N^i_{down,t} >= \frac{N^i_{move,t}}{2}$.

In the first relation, δt is the time step of the simulation and D is the diffusion coefficient of the Fokker-Planck Equation shown in the previous chapter.
The fluctuation in the shape of the order book is tuned by the parameter p and subsequently through the diffusion coefficient D.
Furthermore, unitary measure of probability implies that

$$\delta t \leq \frac{1}{2D}$$

Setting D allows us to choose the desired time-step. In the simulation, $D = 1$.

Finally, we can write the equation for the total number of orders,

$$N^i_t = N^i_{t-\delta t} - N^i_{move,t} + N^{i-1}_{up,t} + N^{i+1}_{down,t}$$

Setting $t = \delta t$ and averaging over many realizations we have

$$< N^i_{\delta t} >= N^i_0 - N^i_0(2D\delta t) + \frac{(2D\delta t)}{2}(N^{i-1}_0 + N^{i+1}_0)$$

The linear initialization of the order book's shape implies (both for bid and ask side)

$$N^{i-1}_0 = N^i_0 - 1$$
$$N^{i+1}_0 = N^i_0 + 1$$

Substituting in the above equation, the last two terms cancel out yielding,

$$< N^i_{\delta t} >= N^i_0$$

This means that the linear shape is stationary $\forall t \in \{0, \delta t, ..., n_{it}\delta t\}$, where $n_{it} \in \mathbb{N}^+$ is the number of iterations.
We are going to treat the reaction and boundary term separately; the above equation holds therefore for

$$i \in \{2, ..., L-1\} \ \setminus \{j\}$$

calling L the order book's size and j, the position at which reaction happens.
The size of the order book in the simulation is $L = 1001$.

Reaction Term

As already mentioned, we want the initial mid-price to be $p_0 = 0$ chosen as a reference frame: for this purpose, *average* available liquidity for the bid and ask side should always be the same, giving the reaction position

$$j = \frac{1}{2}(L+1)$$

At this position, the average number of orders should always be zero, at any time,

$$< N^j_t >= 0 \quad \Rightarrow \quad < p_t >= 0$$

To obtain this, we note that

$$N^j_t = N^j_{t-\delta t} - N^j_{move,t} + N^{j-1}_{up,t} + N^{j+1}_{down,t} - R_t$$

Setting $t = \delta t$ and averaging over many realizations, one has, provided that $N^j_0 = 0$,

$$< N^j_{\delta t} >= N^j_0 - (2D\delta t)N^j_0 + \frac{(2D\delta t)}{2}(N^{j-1}_0 + N^{j+1}_0) - < R_{\delta t} >$$

Giving the linear initialization, $N_0^{j-1} = N_0^{j+1} = 1$, yielding,

$$< N_{\delta t}^{j} >= 2D\delta t - < R >= 0 \quad \Rightarrow \quad < R_{\delta t} >= 2D\delta t$$

This has to be valid for any t, so the random variable R_t has to balance, on average, the *in*-contribution from $j-1$ and $j+1$. Therefore

$$R_t \sim N_{up,t}^{j-1} + N_{down,t}^{j+1}$$

Boundary Terms

In the simulation, ask and bid orders have to diffuse, reacting when they superpose, but leaving in this way, the order book with increasingly less liquidity(number of orders).
Recalling the stationary solution of the Fokker-Planck equation, the latent liquidity was defined as

$$\mathcal{L} = \frac{J}{D}$$

which is, essentially, the order book's slope.
We included the parameter D into the simulation but not J, which can be interpreted as the *average* density of orders injected in the order book; in this way, orders react but liquidity is constantly provided through the parameter J.
Thus, the quantity $J\delta t$ determines the number of orders that, on average, are injected in the book from the boundaries: the orders are taken at each time step, from a Poisson distribution.
Calling $N_{boundary}$, the random variable of the number of orders injected at the boundaries,

$$N_{boundary} \sim \texttt{Poiss}(2J\delta t)$$

The factor 2 comes from the fact that we have two boundaries, one for the ask and one for the bid side.
Poisson distribution can be obtained as a limit case of binomial distribution $\texttt{Bin}(\mathcal{L}, p)$. In fact its distribution average is

$$< N_{boundary} >= \mathcal{L}p = \frac{J}{D}(2D\delta t) = 2J\delta t$$

Another interesting way to look at *Latent Liquidity* is by writing the average ratio of the number of orders injected and reacted in the order book, hence,

$$\frac{< N_{boundary} >}{< R_t >} = \frac{2J\delta t}{2D\delta t} = \frac{J}{D} = \mathcal{L}$$

We can write for the boundaries, setting $t = \delta t$

$$N^1_{\delta t} = N^1_0 - N^1_{move,\delta t} + N_{boundary} + N^1_{down,\delta t}$$

$$N^L_{\delta t} = N^L_0 - N^L_{move,\delta t} + N_{boundary} + N^L_{up,\delta t}$$

Averaging we obtain,

$$< N^1_{\delta t} >=< N^L_{\delta t} >= D\delta t(\mathcal{L} - 1)$$

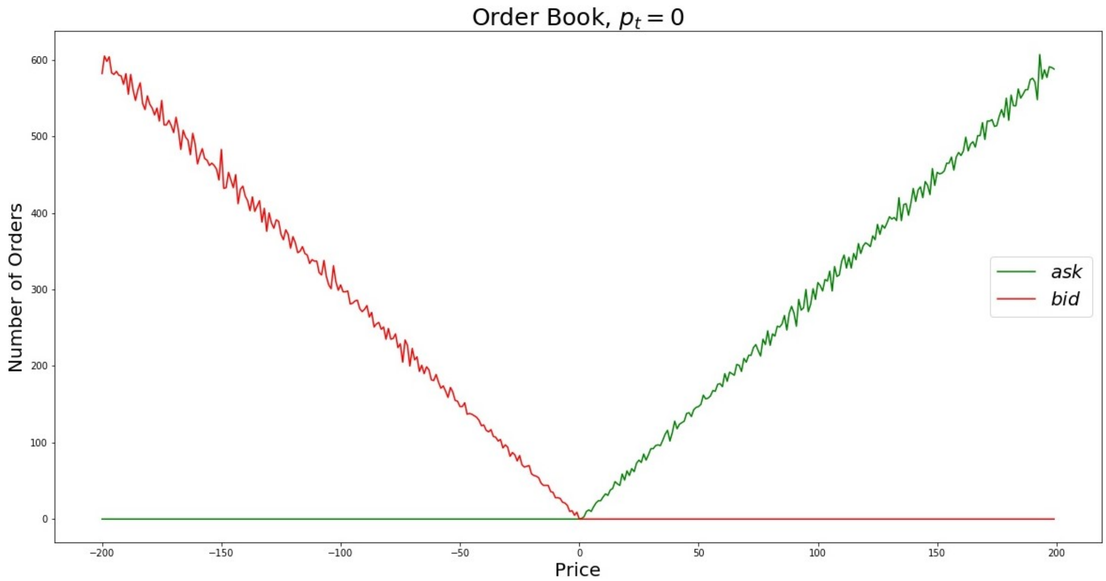

Figure 4.1: A snapshot of the order book's simulation at a random t. One can notice $p_t = 0$. In this case $L = 501$ and $t = 0.1$

4.1.1 Constant Rate m_0

If the meta-orders' execution rate is constant $m_s = m_0$, the integral stochastic equation for the *average* impacted price becomes a bit simpler,

$$y_t - y_0 = \frac{1}{\mathcal{L}} \int_0^t ds \frac{m_0}{\sqrt{4\pi(t-s)}} e^{-\frac{(y_t - y_s)^2}{4D(t-s)}}$$

Neglecting *non-linearities*, hence, $|y_t - y_s|^2 << D(t-s)$, corresponds to having the simulated maximum impacted price, very small with respect to the size of the order book, for instance calling T, the length of the simulation, one would have,

$$< p_T - p_0 >= \mathcal{I}(T) \leq \frac{L}{2}\frac{1}{10}$$

In this way, by pure geometric arguments, we can determine the maximum volume Q, that we can inject as meta-orders. Solving the above integral, we obtain

$$y_T - y_0 = \mathcal{I}(T) \propto 2\frac{m_0}{\mathcal{L}}\sqrt{T}$$

up to a $\sqrt{4\pi}$ constant. Recalling that $D = 1 \Rightarrow \mathcal{L} = J$,

$$2\frac{m_0}{J}\sqrt{T} \leq \frac{L}{2}\frac{1}{10}$$

$$T \leq (\frac{L}{40})^2(\frac{m_0}{J})^{-2}$$

$\frac{m_0}{J}$ is called *participation ratio*: it is a measure of how much orders, on *average*, are injected as metaorders with rate m_0, with respect to *limit orders*, with rate J. Intuitively, an high value of the *participation ratio* would make limit order's trading impossible, *"opening"* the order book and increasing the spread. As we will see, this is very well reproduced by the simulation.

Operatively, at each time step t, orders are injected exactly at the current *mid-price*, p_t. Given that the meta-orders investor has not a particular strategy and m_0 is constant, calling N_{meta}, the random variable of number of meta-orders,

$$N_{meta} \sim \texttt{Bin}(1, m_0\delta t) \quad \Rightarrow \quad < N_{meta} >= m_0\delta t$$

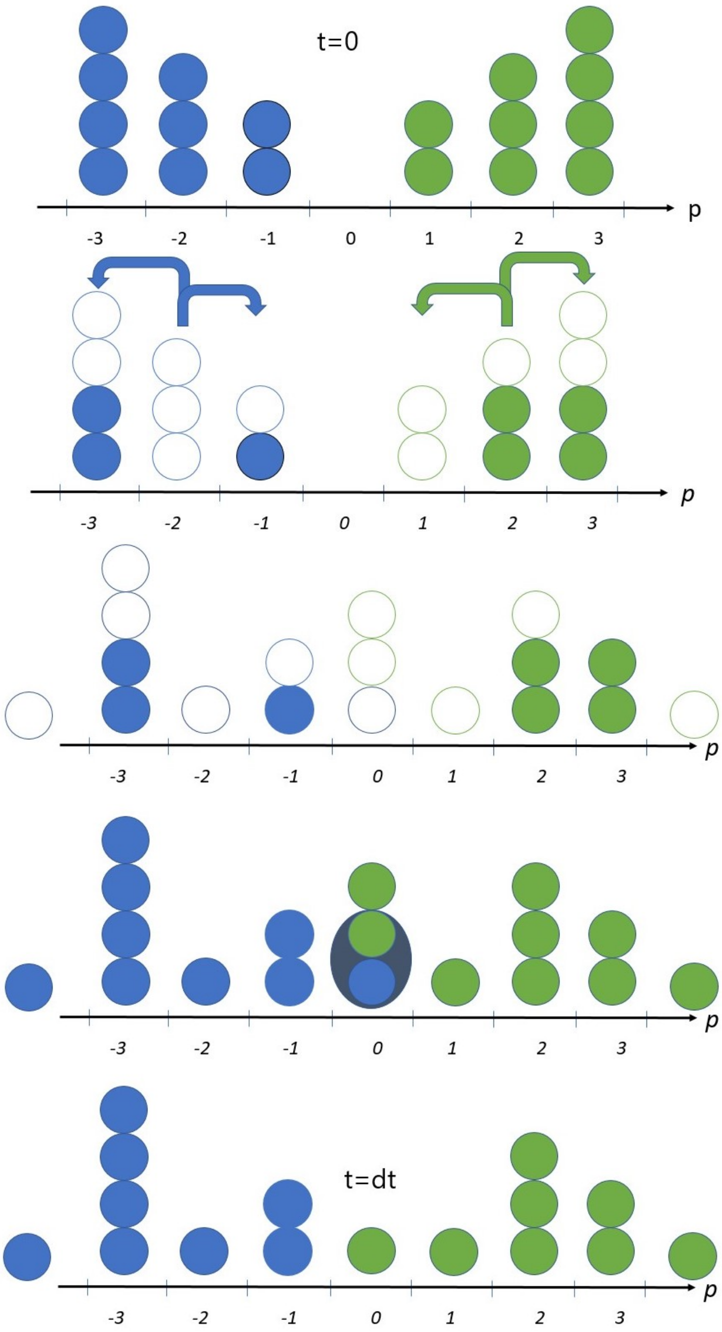

Figure 4.2: [Top to Down]. ***a)*** order book's linear initialization, $t = 0$ and current *mid-price* is $p_0 = 0$. Orders represented as beads: blue for *bid*(`buy`) side and green for the *ask*(`sell`) side. ***b)*** Selected orders are empty beads: they diffuse symmetrically with $p = \frac{1}{2}$. ***c)*** New configuration after diffusion. ***d)*** Beads of different type react one by one when on same price. ***e)*** At $t = \delta t$, after reaction, the new current *mid-price* is $p_{\delta t} = -0.5$, because the *bid* side did not provide enough liquidity.

An important feature that should be reproduce by the simulation is the *square-root impact law*. Recalling the theoretical impact curves for both high and low *participation ratio*

$$\mathcal{I}(Q) \approx \frac{m_0}{J\pi}\sqrt{\frac{Q}{\mathcal{L}}} \qquad for \qquad m_0 << J$$

$$\mathcal{I}(Q) \approx \sqrt{\frac{2Q}{\mathcal{L}}} \qquad for \qquad m_0 >> J$$

Making some substitutions of parameters we can obtain two functions depending on the rescaled time $\frac{t}{T}$, yielding,

$$\mathcal{I}_{th}(t/T) \approx \frac{m_0\sqrt{Q}}{J\pi\sqrt{\mathcal{L}}}\sqrt{\frac{t}{T}} \qquad for \qquad m_0 << J$$

$$\mathcal{I}_{th}(t/T) \approx \sqrt{\frac{2m_0 T}{\mathcal{L}}}\sqrt{\frac{t}{T}} \qquad for \qquad m_0 >> J$$

The 4.3 below shows the evolution of the book compared with the theoretical impact curves.

One can see that in low participation ratio, the theoretical line matches perfectly the *mid-price* curve as expected. The books remains closed and the spread seems roughly constant, on average, throughout the meta-orders' execution and relaxation.
In the high participation ratio, the theoretical formula does not match the mid-price curve, but it matches well the ask side.
The reasons for this may be many: it could be due to discretization factors or a more detailed description could be needed, including the exponential factor in the stochastic integral equation of the equation for the *mid-price*.

In this case, the order book opens itself, increasing the spread and reaching its maximum at $t = T$ and then relaxing a lot slower than the low participation rate case.

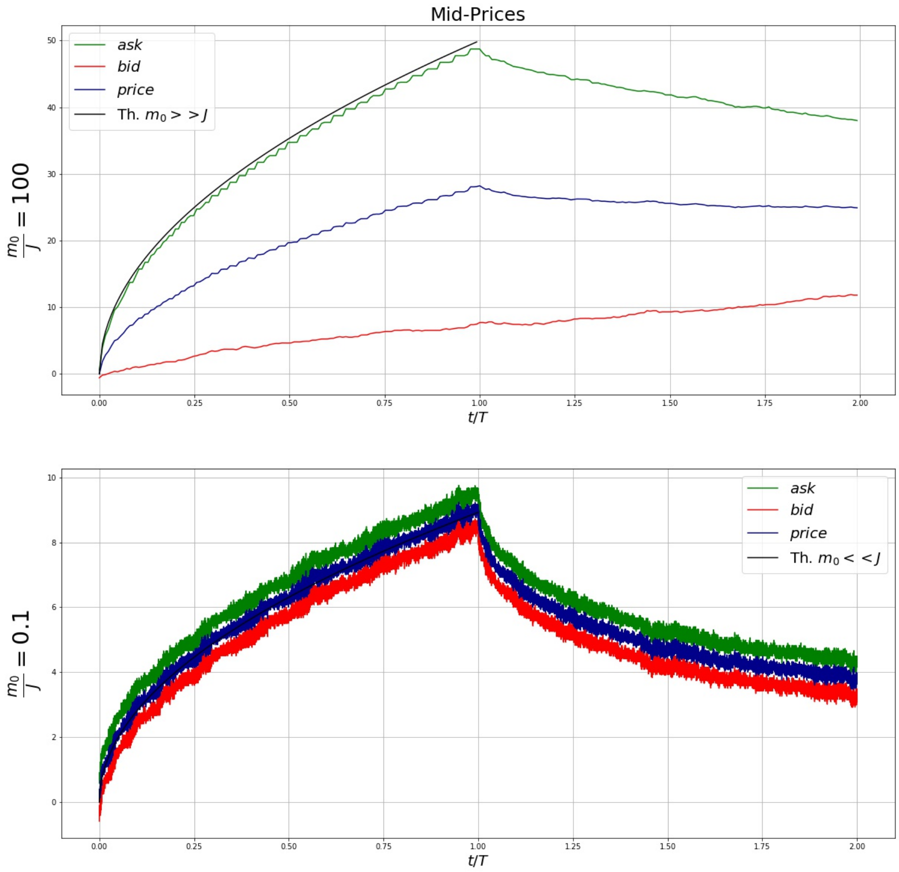

Figure 4.3: Evolution of mid-price p_t, best ask a_t and best bid b_t until $t = 2T$. The theoretical curve is in black for both plots, $\delta t = 0.1$.

4.2 Noise Traders

First, we analyze the behavior of the book solely in the presence of the noise traders. Therefore, $m_0 = 0$ and

$$m_t = \delta t m_t \qquad and \qquad \mathbb{E}[\delta m_t \delta m_{t'}] = \sigma^2 (t - t')^{-\mu}$$

δm_t is an increment of fractional brownian motion (fBn), called fractional Gaussian noise (fGn), with $\mathtt{E}[\delta m_t] = 0$, meaning that there are on average, as many *buy* meta-orders as *sell* meta-orders. The fGn is generated with the *Davies Harte method* for fBm ([**7**], [**8**]).
μ is related to the fractional noise correlation exponent H, through the relation

$$\mu = 2(H - 1)$$

A good way to check the correctness of the simulation is to compute the exponent of the variance against the *Hurst exponent*. The plot 4.4 shows the fitting of the

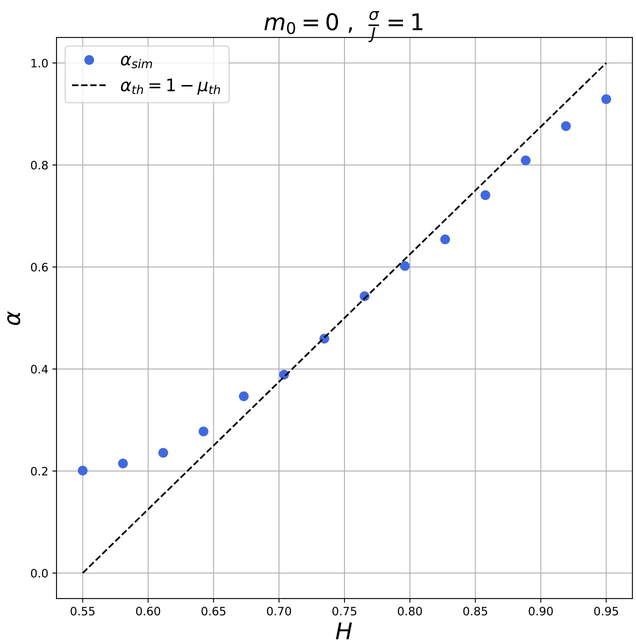

Figure 4.4: Blue dotted: exponent α_{sim} obtained from the fitting. Black dashed: theoretical line $\alpha_{th} = 2(1 - H)$. $\delta t = 0.1$.

mid-price's variance exponent, which we computed previously, $\mathbb{E}[y_t^2] \propto t^{1-\mu}$, where

$\alpha = 1 - \mu$, which is the value plotted versus *Hurst exponent*.
a typical characteristic of the price in financial markets is to be *mean reverting*: this means that the variance exponent has to satisfy the relation,

$$\alpha \leq 1$$

It is worth notice that in case $H = 1$ we have perfect Brownian motion of the mid-price, which is a well known behavior of price returns in financial markets, *market efficiency*. But, from the graph, we can see that in this case, the "experimental" exponent is not exaclty 1.

An important degree of freedom is the parameter σ, amplitude of the noise which can be confronted with the charaterisic current of the order book J, $\frac{\sigma}{J}$.

Tuning the parameter, one finds out that an high value, in particular $\frac{\sigma}{J} > 1$, opens the book symmetrically giving birth to unphysical effects that we have to discard. It is difficult to find out the exact value at which the book opens but it should be roughly one: the threshold value increases with the time-step δt.

Injecting the noise, we introduce also a characteristic timescale in the system that depends on how correlations between meta-orders decay. The other parameter to set is the exponent μ. Empirically, it is known that his $\mu = \frac{1}{2}$ (see [**9**]).
This yields an Hurst exponent of $H = \frac{3}{4} = 0{,}75$.
From now on, we will use this value to run the simulations with additive noise.

The exponent of the variance as showed in the first plot in 4.5, is well reproduced with $\alpha_{sim} - 1 = -\mu_{sim} \approx -\frac{1}{2}$.
For comparison, we report below the same plot as above, with $\frac{\sigma}{J} = 0$. It is immediate to notice that despite being both mean reverting, the presence of the correlated meta-orders imply longer range correlations in the price; It is "less" *mean-reverting*.

The effect of the noise has no transient state, if switched on at a certain time, the effect on the spread $< s_t > = < a_t > - < b_t >$, is almost instantaneous and shifts up the the value of the spread, that remains stationary hereafter, (4.6).

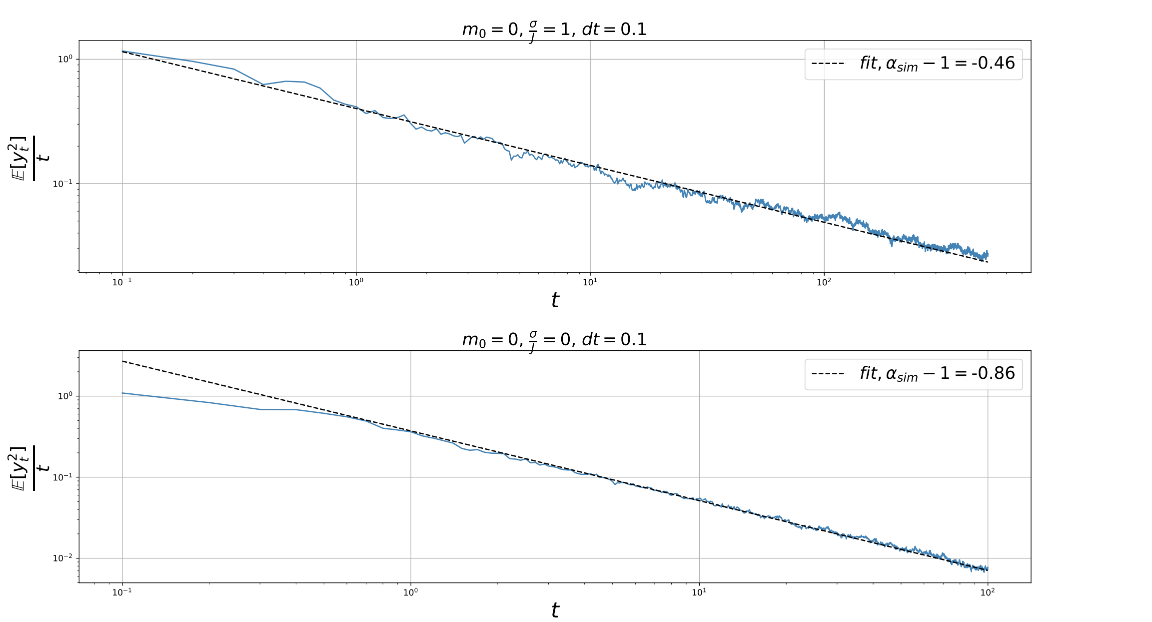

Figure 4.5: In blue: signature plot of the price y_t in log-log scale. In Black dashed: slope line fitted from data with exponent $\alpha_{sim} - 1$

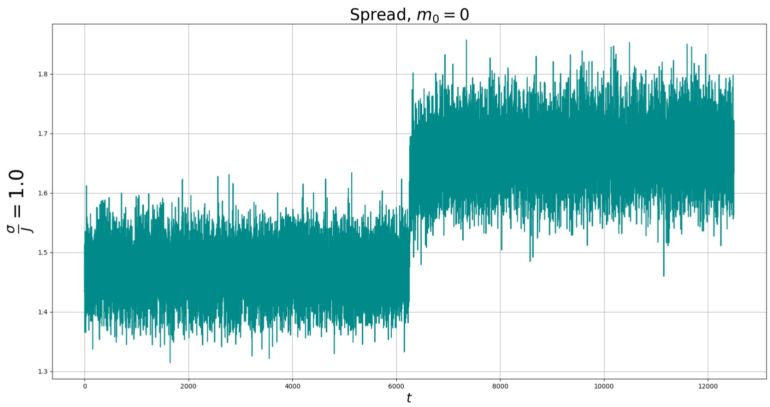

Figure 4.6: Average spread. $\delta t = 0.5$. The noise starts at half number of steps n_{it}

4.3 Herding Behaviour

Giving the general form of meta-orders' rate,

$$m_t = m_0 + \delta m_t$$

we want to investigate if the impact curve changes qualitatively with respect to the noiseless case. From the model's solution we recall the stochastic equation of the price in the linear approximation, hence,

$$y_t = \frac{1}{\mathcal{L}} \int_0^t ds \frac{m_0}{\sqrt{4\pi(t-s)}} + \frac{1}{\mathcal{L}} \int_0^t ds \frac{\delta m_s}{\sqrt{4\pi(t-s)}}$$

Applying the expected value with respect to noise's realizations, given that $\mathbb{E}[\delta m_s] = 0$, one has

$$\mathbb{E}[y_t] = \frac{1}{\mathcal{L}} \int_0^t ds \frac{m_0}{\sqrt{4\pi(t-s)}}$$

The impact behavior on the price in presence of symmetric noise traders should not depend significantly on the long-range correlations of meta-orders, if the noise does not dominate the dynamics, with respect to the characteristic dynamics of the order book.
We choose to work therefore in the regime $\frac{\sigma}{J} = 1$. That's because we are sure the simulation reproduces correctly the theoretical variance exponent of the price.

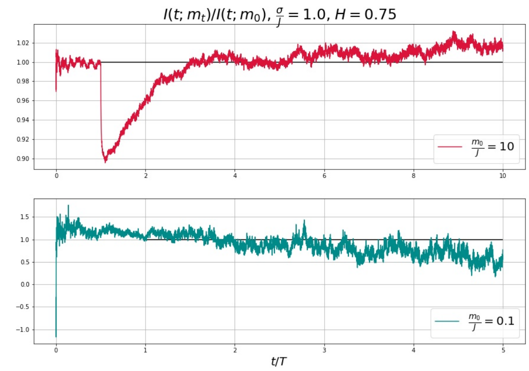

Figure 4.7: Impact ratio between the general case and the constant rate case. $\delta t = 0.5$

In the 4.7 the ratio $\frac{I(t;m_t)}{I(t;m_0)}$ proves that meta-orders noise traders do not change significantly, aside from fluctuations, the behavior of the impact curve, as predicted by the theory. This is particularly true if $m_0 >> J$. In low participation ratio

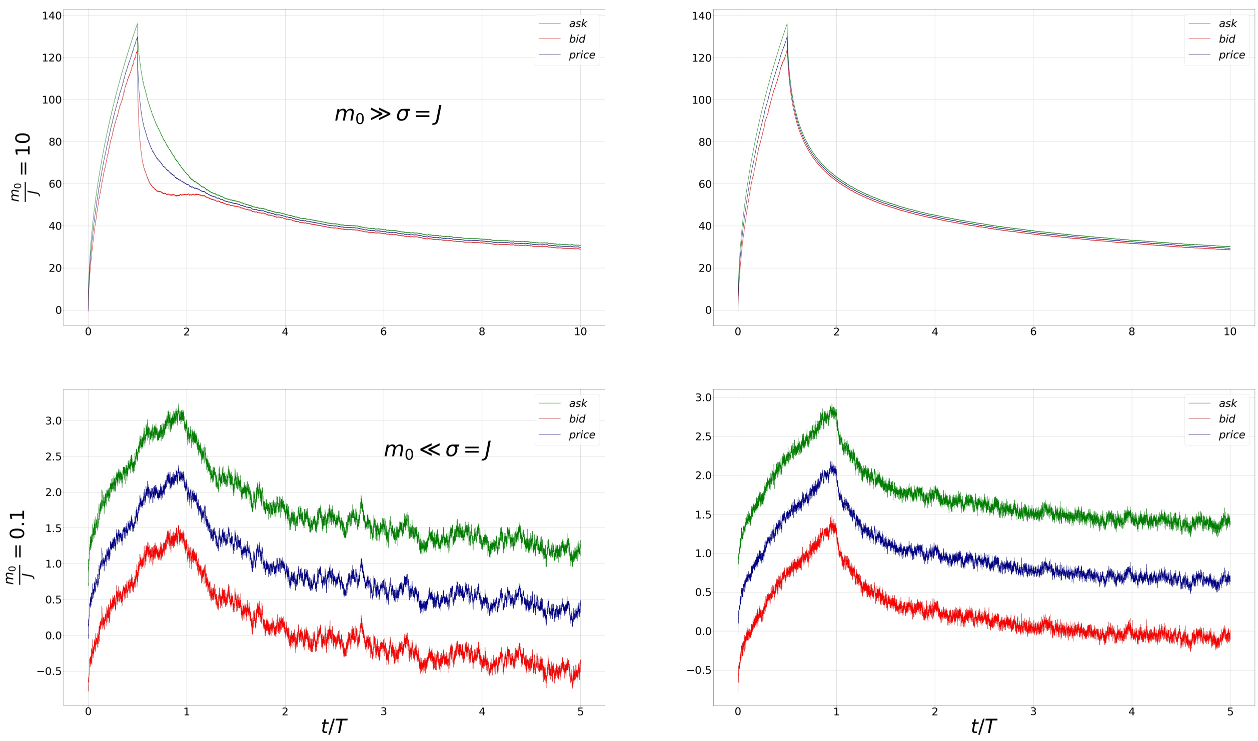

Figure 4.8: Mid-prices for $\frac{\sigma}{J} = 1$ (left side) and $\frac{\sigma}{J} = 0$ (right side). $\delta t = 0.5$

instead there looks to be a little imbalance between the two: the ratio stays slightly over 1 until the end of the execution, but that could be a discretization problem, since the number of orders executed is significantly lower. At $\frac{t}{T} = 1$ something interesting happens. The relaxation looks qualitatively different with respect to the noiseless case, at least in high participation ratio.

This fact can be better seen in 4.8 in the top-left figure: we are in high participation ratio with respect to both J and σ.

After the end of the execution, in fact, the presence of the noise creates an imbalance in the order book, in this case in the bid size, pushing down the the best price b_t. In that region, during relaxation, orders react faster than they can diffuse, opening the book and increasing the spread. The effect lasts roughly twice the execution time T, closing back the book and returning to the *non-impacted* price, always slower than the noiseless case.

In the low participation ratio instead, one cannot appreciate any substantial difference, apart from relaxation which tends to be faster, as one can see better in 4.7, and the average signals are less noisier in the noiseless case.
The advantage of this strategy, for the constant rate investor, is being able to be *"less visible"* to other traders, and therefore, disclosing less information, essential when trying to make profit. This requires also a longer. execution-time.
Instead, in the previous case $m_0 >> \sigma = J$, the rate is too high and the price is pushed-up, not convenient for the buyer on a first instance, but useful if we want to sell after at a smaller rate $_0$, hoping to make a profit.
Furthermore, the log-log plot of the Impact below shows in detail that the *square-root law* is reproduced (4.9).

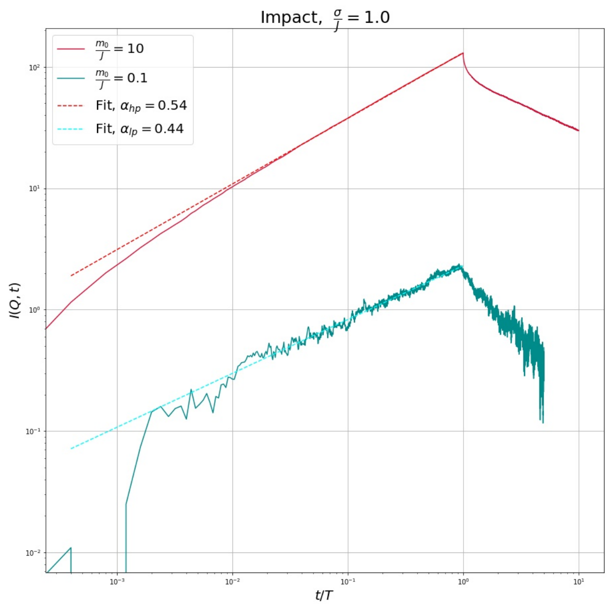

Figure 4.9: log-log Impact for high and low participation ratio. $\delta t = 0.5$

In the next chapter we will, show that the *square-root* hold also in Q, volume of *meta-orders* traded. Indeed, it can be showed that Q is a function of time through the relation,

$$Q(t) = m_0 t$$

4.4 Co-Impact

In a recent paper [**10**], it was investigated empirically the effect of meta-orders on the square-root law. In particular, it was found that, although the data is very noisy, there might be some cross-over from *square-root* to *quasi-linear*, ending up with a plateau.
The explanation given, intuitively, was that, if we trade a fraction of the total volume traded in some unit time, ϕ, which can be mapped onto the constant rate parameter m_0, and to we add another fraction ϕ_m with the same sign(buy or sell), since the impact function is additive, one should find

$$\mathcal{I}(\phi + \phi_m) = C\sqrt{\phi + \phi_m}$$

that behaves linearly and has a plateau if $\phi << \phi_m$. Given that we don't have any trend in the noise meta-orders, since $\mathbb{E}[\delta m_s] = 0$, we should not find any plateau, for ϕ tending to zero. It is worth trying though, if the square-root law is kept and if the there are some similarities between the curve fitted from empirical data and the plots obtained from the simulation. The idea is to repeat the same the same data analysis done in the above mentioned paper, with simulation data. We collect simulation's price data for 80 different participation ratios, with $n_{it} = 1000$

$$10^{-3} \leq \frac{m_0}{J} \leq 10^2$$

computing , each time, the parameter ϕ defined as:

$$\phi = \frac{1}{1 + \frac{V_{react}}{Q_0} + \frac{\sigma}{m_0 \delta t} < Q_{noise} >}$$

Thereafter, the data has been binned in order not to lose the information on the variance. The result is good for high values of ϕ: the fluctuations are not dominant since volume traded is big enough. Nevertheless, it looks very noisy if the quantity traded at constant rate is small(4.10). Finally, the 4.11 shows the fitted slope which gives value $\delta_{fit} = 0.44$. Empirically the parameter lays between 0.4 and 0.7, therefore it seems an acceptable result.

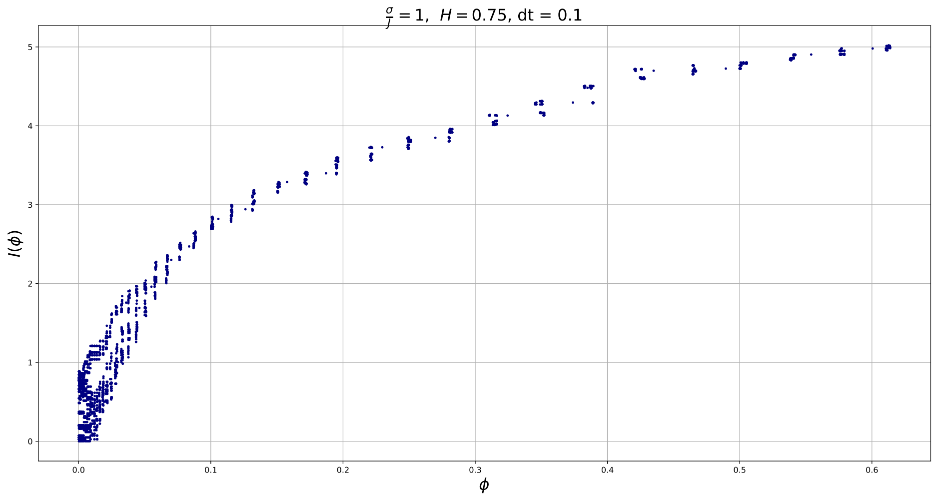

Figure 4.10: Co-Impact vs. Participation ratio ϕ

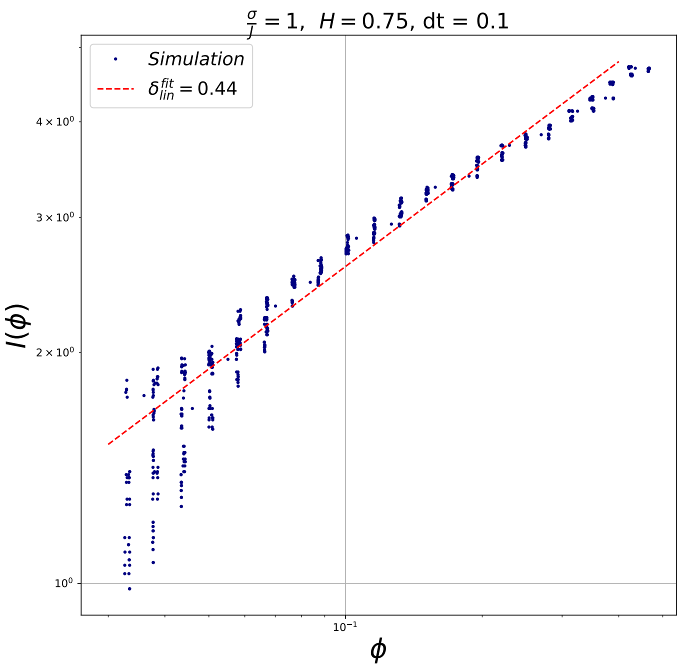

Figure 4.11: In dark blue: log-log Impact vs. participation ratio ϕ. In red: fitted slope from data.

4.5 Concluding Remarks

Overall, the simulation demonstrate a quite satisfactory result. On one hand, the code seems robust, and works well in the regimes suggested by the recent empirical discoveries.

On the other hand, at the moment, the simulation is not able to reproduce a clear cross-over between square-root and linear behavior.

it may require further investigation, in particular, adding some new features that can help confronting with the reality of financial markets. The code seems to be stable and robust with respect the *Hurst exponent*, which can allow to explore different time correlation length of *meta-orders*.
Also the square-root law is well reproduced.

There seems to space for future investigation. It could be worth trying to put in place a trend in the noisy meta-orders,

$$\mathtt{E}[\delta m_t] \neq 0$$

If there exists a plateau, it should be proportional to the square root of participation ratio given the trend, times a constant.
It could also depend on the amplitude of the fGn, σ.
4.12 shows the empirical plot. The cross-over that appears to be present from square-root to an almost linear behavior. There seems to be a sort of plateau.

Indeed, the fitting function is a square-root plus constant B.

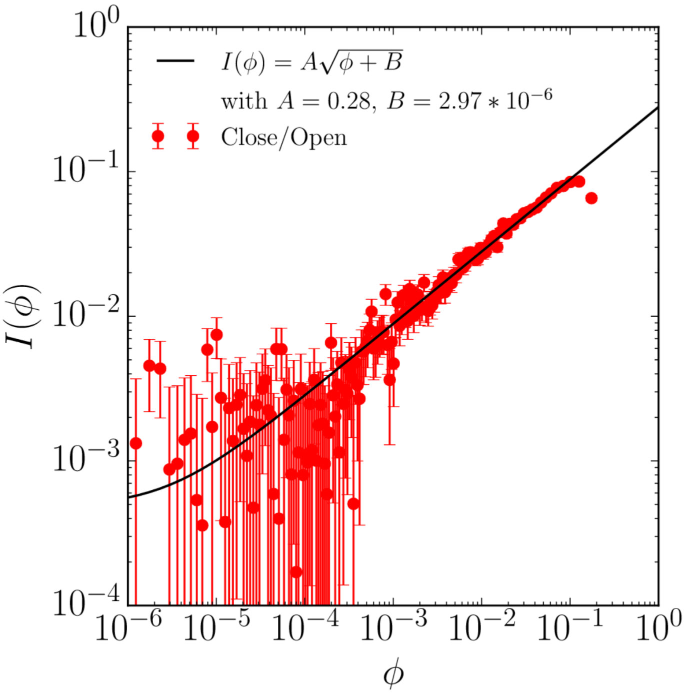

Figure 4.12: Empirical Co-Impact plot showing clearly the two regimes and a sort of plateau. Graph taken from [**10**]. *Data taken from ANCERNO database.*

Bibliography

[1] R.Almgren, C. Thum, E. Hauptmann, H. Li, *Direct estimation of equity market impact*, Risk, 18(7), 5762 (2005).

[2] B. Tóth, Y. Lempérière, C. Deremble, J. De Lataillade, J. Kockelkoren, J.-P. Bouchaud, *Anomalous price impact and the critical nature of liquidity in financial markets*, Phyisical Review X, 1(2), 021006 (2011).

[3] D. M. Cutler, J.M. Poterba, and L. H. Summers, Journal Portfolio Management 15, 4 (1989).

[4] A. Joulin, A. Lefevre, D. Grunberg, and J.-P. Bouchaud, Wilmott Magazine Sep, 1 (2008).

[5] J. Donier, J. Bonart, I. Mastromatteo, and J.-P. Bouchaud, *A fully consistent, minimal model for non-linear market impact*, arXiv:1412.0141v4 (2015).

[6] F. Bucci, M. Benzaquen, F. Lillo, and J.-P. Bouchaud, *Crossover from Linear to Square-Root Market Impact*, arXiv:1811.05230v1 (2018).

[7] R.B. Davies, D.S. Harte, Tests for Hurst effect, Biometrika, 74 (1987).

[8] Ton Dieker, Simulation of fractional Brownian motion, University of Twente, Department of Mathematical Sciences, Amsterdam (2004).

[9] J.-P. Bouchaud, J. Donier, J. Bonart, and M. Gould, Trades, quotes and price: Financial Markets under the microscope, Cambridge University Press (2018).

[10] F. Bucci, I. Mastromatteo, Z. Eisler, F. Lillo, J.-P. Bouchaud, and C.-A. Lehalle, Co-impact: Crowding Effects in Institutional Trading Activity, arXiv:1084.09565v2 (2018).

[11] L. Dall'Amico, A. Fosset, J-P. Bouchaud, M. Benzaquen, How does liquidity get revealed in the limit order book, arXiv:1808.09677v2 (2018).