UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia "Galileo Galilei"

Corso di Laurea in Fisica

Tesi di Laurea

Diffusione anomala di polimeri critici e dinamica dei mercati finanziari

Anomalous diffusion of critical polymers and financial markets dynamics

Relatore	Laureando
Prof. Fulvio Baldovin	Alberto Raiola

Anno Accademico 2021/2022

Contents

1 Polymer Physics

1.1 Introduction and aim of the thesis

The diffusion of a single monomer immersed in a fluid can be described using Brownian motion and it is well known that the distribution of its displacements is Gaussian. Instead, as we will show below, the distribution of a polymer center of mass may follow a sharp peaked, heavy tailed non-Gaussian distribution. Indeed, one of the current research topics in Polymer Physics is the Brownian non-Gaussian diffusion and this particular behaviour is modeled in terms of a microscopic process called polymerization/depolymerization which is mathematically described using tools provided by the queueing theory.

Distributions with the same anomalous features also appear in another context, financial markets. In particular a non-Gaussian behaviour is observed in the distribution of assets returns, that is the logarithm of the assets price variation. Price trends depend on the number of orders that traders make and they are collected in a list called order book.

Aim of this thesis is to make a link between these two worlds: Polymer Physics and financial markets dynamics. In particular, queueing theory will be the key tool to reach this scope. In this section the first world is described.

1.2 The diffusion of single particles: Brownian motion

Before describing polymer center of mass dynamics, it is propedeutic to show the behaviour of single particles immerse in a fluid.

Starting from a one-dimensional fluid, let $n(x,t)$ and $p_X(x,t)dx = n(x,t)dx/N$ be respectively the local concentration of particles and the probability of finding a particle between x and $x+dx$ at time t. Due to the random collisions with the molecules of the solvent one can imagine that gradients in the concentration can act as general forces inducing a matter flow. Fick's first law relates this diffusive flow $j(x,t)$ to the gradient of concentration:

$$j(x,t) = -D_0 \frac{\partial n(x,t)}{\partial x}\,, \tag{1}$$

where D_0 is the diffusive constant with the dimension of a length squared divided by time. Now, as a consequence of the mass conservation, the following continuity equation holds:

$$\frac{\partial n(x,t)}{\partial t} = -\frac{\partial j(x,t)}{\partial x} \tag{2}$$

and putting the expression for the diffusive flow obtained in Eq. (1) in the previous we obtain the diffusion equation:

$$\frac{\partial n(x,t)}{\partial t} = D_0 \frac{\partial^2 n(x,t)}{\partial x^2}\,. \tag{3}$$

Solving with initial condition $p_X(x,0) = n(x,0)/N \propto \delta(x)$ one obtains the following Gaussian distribution for a Brownian particle position:

$$p_X(x,t) = \frac{1}{\sqrt{4\pi D_0 t}} e^{-\frac{x^2}{4D_0 t}} \tag{4}$$

with variance $2D_0 t$.

The Brownian particle position $X(t)$ is then a stochastic process and it can be well described using the Langevin approach. Let us consider a one-dimensional particle immersed in a fluid whose viscosity is indicated by η and make the hypothesis that the net force exerted on the particle at time t by the surrounding fluid molecules can be resolved in two components:

- a dissipative drag force $-\xi V(t)$ where $V(t)$ is the particle velocity at time t and $\xi \propto \eta/m$ is the friction constant;
- a randomly fluctuating force $F(t)$ which follows these assumptions: $\mathbb{E}[F(t)] = 0$ (zero average), it is temporally uncorrelated, that is $\mathbb{E}[F(t')F(t'')] = C\delta(t' - t'')$, it is assumed to be statistically independent of $V(\tau)$ for all $\tau < t$.

The equations of motion are obtained applying Newton's second law:

$$\begin{cases} \frac{dX}{dt} = V(t) \\ \frac{dV}{dt} = -\xi V(t) + \frac{F(t)}{m} - \frac{1}{m}\frac{d\Phi}{dx} \end{cases} , \tag{5}$$

where the Brownian particle mass is m and Φ is an external potential.
Considering an overdumped situation (more rigorously, the Smoluchowski time scale) the previous becomes:

$$\frac{dX}{dt} = \frac{1}{\xi m}F(t) + \frac{1}{m}\frac{d\Phi}{dx} \tag{6}$$

and ignoring the external field, we can integrate with initial condition $X(0) = 0$:

$$X(t) = \int_0^t \frac{1}{\xi m}F(\tau)d\tau\,. \tag{7}$$

Applying the hypothesis on $F(t)$:

$$\mathbb{E}[X(t)] = \int_0^t \frac{1}{\xi m}\mathbb{E}[F(\tau)]d\tau = 0\,; \tag{8}$$

$$\begin{aligned} \mathbb{E}[X(t')X(t'')] = \int_0^{t'} d\tau' \int_0^{t''} d\tau'' \frac{1}{\xi^2 m^2}\mathbb{E}[F(\tau')F(\tau'')] = \\ = \int_0^{t'} d\tau' \int_0^{t''} d\tau'' \frac{1}{\xi^2 m^2} C\delta(\tau' - \tau'') = \\ = \frac{C}{\xi^2 m^2}\int_0^{t'} d\tau' \int_0^{\infty} d\tau'' H(t'' - \tau'')\delta(\tau' - \tau'') = \\ = \frac{C}{\xi^2 m^2}\int_0^{\infty} d\tau' H(t' - \tau')H(t'' - \tau') = \frac{C}{\xi^2 m^2}t'\,, \end{aligned} \tag{9}$$

where $t' < t''$ and H is the Heaviside step function. Then:

$$\mathbb{E}[X^2(t)] - \mathbb{E}[X(t)]^2 = \mathbb{E}[X^2(t)] = 2D_0 t\,, \tag{10}$$

where $2D_0 = C/\xi^2 m^2$, as expected from Eq. (4).
We can rewrite Eq. (7) as:

$$X(t) = \sqrt{2D_0}B(t)\,, \tag{11}$$

where B is named Wiener process and it is subjected to the following conditions: $B(0) = 0$, $\mathbb{E}[B(t)] = 0$, $\mathbb{E}[B(t')B(t'')] = t'$ if $t' < t''$. Ignoring again the external field, Langevin equation (6) can be rewritten as:

$$dX(t) = \sqrt{2D_0}dB(dt)\,. \tag{12}$$

1.3 The diffusion of polymers: Rouse model

Polymers are long chain molecules composed of many repeating subunits called monomers. Polystyrene is an example of synthetic polymer; proteins and DNA have a natural origin. Polymers have very different chemical-physical properties and behaviours from each other and these are determined by the characteristics of the monomers and their mutual binding. In the context of this work we are not interested in local properties but in more general ones emerging from the fact that polymers are just long chains and from thermodynamic quantities such as the quality of the solvent in which the polymers are immersed [3].

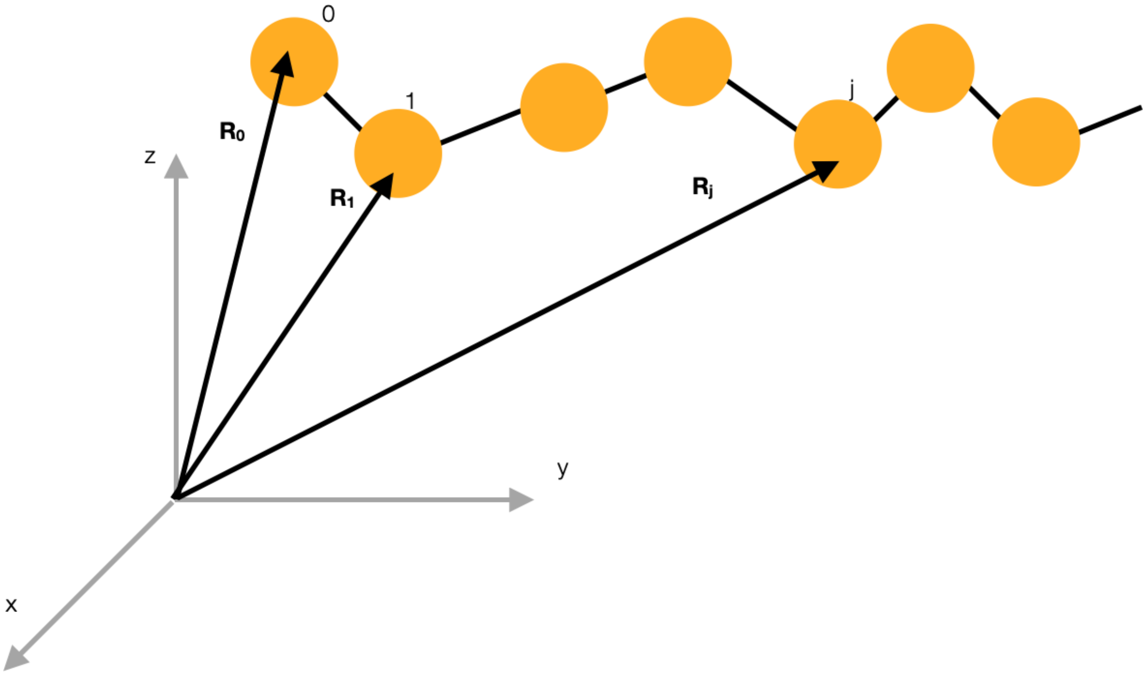

Figure 1: In this picture we show a polymer: orange beads represent monomers, grey arrows represent a Cartesian frame of reference and black arrows are the monomers position vectors $\mathbf{R}_j$.

The conformation of a 3D polymer may be described giving the position vector of its monomers. If the chain consists of N beads then there are N position vectors $\mathbf{R}_0, \mathbf{R}_1, \ldots, \mathbf{R}_{N-1}$. and if the monomers have the same mass, the position of the polymer center of mass $\mathbf{R_G}$ can be computed as follow:

$$\mathbf{R_G} = \frac{1}{N} \sum_{j=0}^{n-1} \mathbf{R_j}. \tag{13}$$

Rouse model is used in polymer physics to describe the dynamics of an ideal chain moving through a liquid [2]. In this model the single chain diffusion is due to the Brownian motion of every single bead composing it. So, every bead is subjected to two forces: a drag one and a thermally stochastic one. The motion of a bead through the liquid induce a velocity field in the liquid which will be felt by all the other beads. In the Rouse model, these hydrodynamic interactions are neglected. Also we assume there are not excluded volume interactions between the beads and we neglected the interactions between non-consecutive beads: the interaction between two consecutive beads is represented by an harmonic spring. All the springs of the chain are assumed to have the same elasticity constant k which can be estimate from the quadratic average position of a monomer $\mathbb{E}\left[\mathbf{R}_j \cdot \mathbf{R}_j\right] = b^2$. In fact, applying equipartition theorem:

$$\mathbb{E}\left[\frac{1}{2}k\mathbf{R}_j^2\right] = \frac{1}{2}kb^2 = \frac{3}{2\beta} \to k = \frac{3}{\beta b^2}\,, \tag{14}$$

where $\beta = 1/(k_B T)$, k_B is the Boltzmann constant and T is the absolute temperature. Using this model it is possible to find the diffusion coefficient of the polymer center or mass. The total force applied to the $j-$th bead is:

$$\mathbf{F}_j = \frac{3}{\beta b^2}(2\mathbf{R}_j - \mathbf{R}_{j-1} - \mathbf{R}_{j+1})\,. \tag{15}$$

Remembering the Langevin equation expressed in the Smoluchowski time scale (Eq. (6)):

$$\frac{d\mathbf{R}_j}{dt} = -\frac{3}{\beta\gamma b^2}(2\mathbf{R}_j - \mathbf{R}_{j-1} - \mathbf{R}_{j+1}) + \mathbf{f}_j\,, \tag{16}$$

where $\gamma = m\xi$ and $\mathbf{f}_j = \mathbf{F}_j/\xi m$ and applying the same reasoning to the other beads we obtain the following equations of motion:

$$\begin{cases} \frac{d\mathbf{R}_0}{dt} = -\frac{3}{\beta\gamma b^2}(\mathbf{R}_0 - \mathbf{R}_1) + \mathbf{f}_0 \\ \frac{d\mathbf{R}_j}{dt} = -\frac{3}{\beta\gamma b^2}(2\mathbf{R}_j - \mathbf{R}_{j-1} - \mathbf{R}_{j+1}) + \mathbf{f}_j \qquad 1 \le j \le N-2\,. \\ \frac{d\mathbf{R}_{n-1}}{dt} = -\frac{3}{\beta\gamma b^2}(\mathbf{R}_{N-1} - \mathbf{R}_{N-2}) + \mathbf{f}_{N-1} \end{cases} \tag{17}$$

Now if we calculate the time derivative of the polymer center of mass we obtain:

$$\frac{d\mathbf{R}_G}{dt} = \frac{1}{N}\sum_{j=0}^{N-1}\mathbf{f}_j\,. \tag{18}$$

Integrating:

$$\mathbf{R}_G(t) = \mathbf{R}_G(0) + \int_0^t d\tau \frac{1}{N}\sum_j \mathbf{f}_j(\tau) \tag{19}$$

and averaging:

$$\mathbb{E}\left[\mathbf{R_G}(t) - \mathbf{R_G}(0))^2\right] = \mathbb{E}\left[\int_0^t d\tau \int_0^{t'} d\tau' \left(\frac{1}{N}\sum_j \mathbf{f}_j(\tau)\right)\left(\frac{1}{N}\sum_k \mathbf{f}_k(\tau')\right)\right] = \frac{6D_0}{N}t\,. \tag{20}$$

This result is crucial: Rouse model states the diffusion coefficient of the polymer center of mass depends on the size of the polymer:

$$D(N) = \frac{D_0}{N^\alpha}\,, \tag{21}$$

where $\alpha = 1$. In other models the parameter α is different. In the Zimm model, in which one considers hydrodynamic interactions between monomers, one has $\alpha = 1/2$ while in the reptation model $\alpha = 2$ [4]. Using the Wiener process notation, we can state that $\mathbf{R}_G$ behaves as:

$$d\mathbf{R_G}(t) = \sqrt{2D(N)}d\mathbf{B}(dt)\,. \tag{22}$$

In the following, we will consider only one-dimensional polymers.

1.4 Critical polymers in the mean field limit and birth and death processes

Now we relax the size-fixed hypothesis and study chemostatted polymers immersed in a monomers bath. In the polymerization process, monomers join the chain and in the depolymerization process they leave the polymer chain: this imply the non-constancy of the polymer size. Such a situation can be well described using the grand canonical ensemble formalism, where the monomer fugacity z governs the partition function $\mathcal{Z}_{gc}$ and the equilibrium distribution $p^*_N(n)$. Let z_c be the critical fugacity, $g = z/z_c$ and γ the universal entropic exponent. It can be shown [4] that, close to criticality, the grand canonical partition function and the equilibrium distribution behave asymptotically as:

$$\mathcal{Z}_{gc}(z) = \sum_{n=1}^{\infty} (z/z_c)^n n^{\gamma-1} \quad \text{and} \quad p^*_N(n) = \frac{(z/z_c)^n n^{\gamma-1}}{\mathcal{Z}_{gc}(z)} \,. \tag{23}$$

In the mean field limit $\gamma \to 1$ [4] we obtain:

$$\mathcal{Z}_{gc}(z) = \frac{g}{1-g} \quad \text{and} \quad p^*_N(n) = g^{n-1}(1-g) \,. \tag{24}$$

The average length and variance of the polymer size are:

$$\mathbb{E}[N] = \frac{1}{1-g} \quad \text{and} \quad \sigma^2_N = \frac{g}{(1-g)^2} \,. \tag{25}$$

Here it can be appreciated the meaning of the critical fugacity: $g = 1$ separates a $g < 1$ dilute phase of finite average size and variance from a $g > 1$ phase of infinite growth. In the limit $g \to 1^-$ we have $\sigma^2_n = g\mathbb{E}[N]^2 \sim \mathbb{E}[N]^2$ and this shows that near the critical point non-normal fluctuations arise.

Another powerful way to describe chemostatted polymers in the mean field limit is thinking polymers of as queues of monomers. If the symbol A_N represents a N-sized chain, it can be thought the polymer undergoes a chemical reaction like:

$$A_N + A_1 \leftrightarrow A_{N+1} \,.$$

The mathematical discipline studying such queueing systems is named queueing theory.

All queuing systems can be broken down into individual sub-systems consisting of entities queuing for some activity and they can be thoughts of as customers waiting queueing for service. The simplest of this systems are single queues, also called waiting lines. Single queues could be also connected to create a queues network but in this work we are not interested in such complex systems. As we know queues are a common every-day experience: typical examples might be waiting for service in banks or supermarkets or waiting for a train or a bus.

The behaviour of a queue can be described by a birth-death process. They are stochastic processes in which jumps from a state are only allowed to neighbouring states: a jump to the right represents birth while a jump to the left represents death.

If $N(t) \in \mathbb{N}^*$ is a population size then a birth-death process is a Markow process $\{N(t), t \geq 0\}$ such that in an interval $(t, t+\Delta t)$ each individual in the population has the probability $\lambda\Delta t + o(\Delta t)$ of giving birth to a new individual and the probability $\mu\Delta t + o(\Delta t)$ of dying. λ and μ are called birth rate and death rate respectively and generally they are size-dependent but in the following we will consider them size independent.

Fixing the ratio $\lambda/\mu = g$, the state probabilities $p_N(n,t)$ of a chemostatted polymer size being in state n at time t in the mean field limit can be described by the master equation:

$$\begin{cases} \frac{d}{dt}p_N(1,t) = -\lambda p_N(1,t) + \mu p_N(2,t) \\ \frac{d}{dt}p_N(n,t) = \lambda p_N(n-1,t) + \mu p_N(n+1,t) - (\lambda+\mu)p_N(n,t) \quad n > 1 \end{cases} \tag{26}$$

with initial condition $n = n_0$ at time $t_0 = 0$ and size independent birth and death rate [4]. In particular we can write $\lambda = g\mu$ where μ is determined by the chemistry of the system. In this case Eq. (26) describes an M/M/1 queue [4] and represents the queue length in a system having a single server where arrivals are determined by a Poisson process and job service times have an exponential distribution.

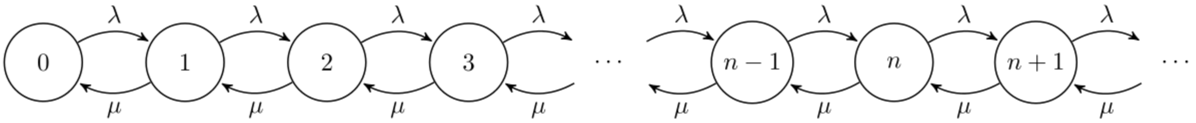

Figure 2: A schematic illustration of a M/M/1 process

It is interesting to solve these equations in the stationary case, that is when all the derivatives in Eq. (26) are zero ($d(p_N(n,t))/dt = 0\ \forall n$). Let us define $p^*_N(n)$ the stationary probability function. Starting from $n = 1$ we obtain the following identity:

$$0 = -\lambda p^*_N(1) + \mu p^*_N(2) \tag{27}$$

and then

$$p^*_N(2) = \frac{\lambda}{\mu} p^*_N(1) = g p^*_N(1)\,. \tag{28}$$

From the second equation of the system Eq. (26) we have:

$$0 = \lambda p^*_N(1) + \mu p^*_N(3) - (\lambda + \mu) p^*_N(2) \tag{29}$$

and then:

$$p^*_N(3) = \left(1 + \frac{\lambda}{\mu}\right) p^*_N(2) - \frac{\lambda}{\mu} p^*_N(1) = g^2 p^*_N(1)\,. \tag{30}$$

By induction it is easy to obtain:

$$p^*_N(n) = g^{n-1} p^*_N(1)\,. \tag{31}$$

Imposing the normalization condition $\sum_n p^*_N(n) = 1$ we obtain the following condition:

$$\sum_{n=1}^{\infty} g^{n-1} p^*_N(1) = 1\,. \tag{32}$$

If $g < 1$, remembering the geometric series properties we obtain as expected the same stationary solution of Eq. (24).
There exists a time-dependent solution of Eq. (26) [4]:

$$p_N(n,t) = e^{-(\lambda+\mu)t} \left[g^{\frac{n-n_0-2}{2}} I_{n-n_0-2}(at) + g^{\frac{n-n_0-3}{2}} I_{n+n_0-1}(at) + (1-g) g^{n-1} \sum_{j=n+n_0}^{\infty} g^{-j/2} I_j(at) \right], \tag{33}$$

where $a = 2\sqrt{\lambda\mu}$ and I_n is the modified Bessel function of the first kind. It can be shown [4] that, in the limit $t \to +\infty$, this solution reaches the stationary solution $p^*_N(n)$ computed before. This is shown in the following graph. From Eq. (33) it is possible [4] to define the time scale τ of the exponential decay of the auto-correlation coefficient $r(t) \approx e^{-t/\tau}$:

$$\tau = \frac{1+g}{(1-g)^2 \mu}\,. \tag{34}$$

If $t \ll \tau$ we have $p_N(n,t) \sim \delta_{n,n_0}$ and if $t \gg \tau$ we have $p_N(n,t) \sim p_N^*(n)$ [4].

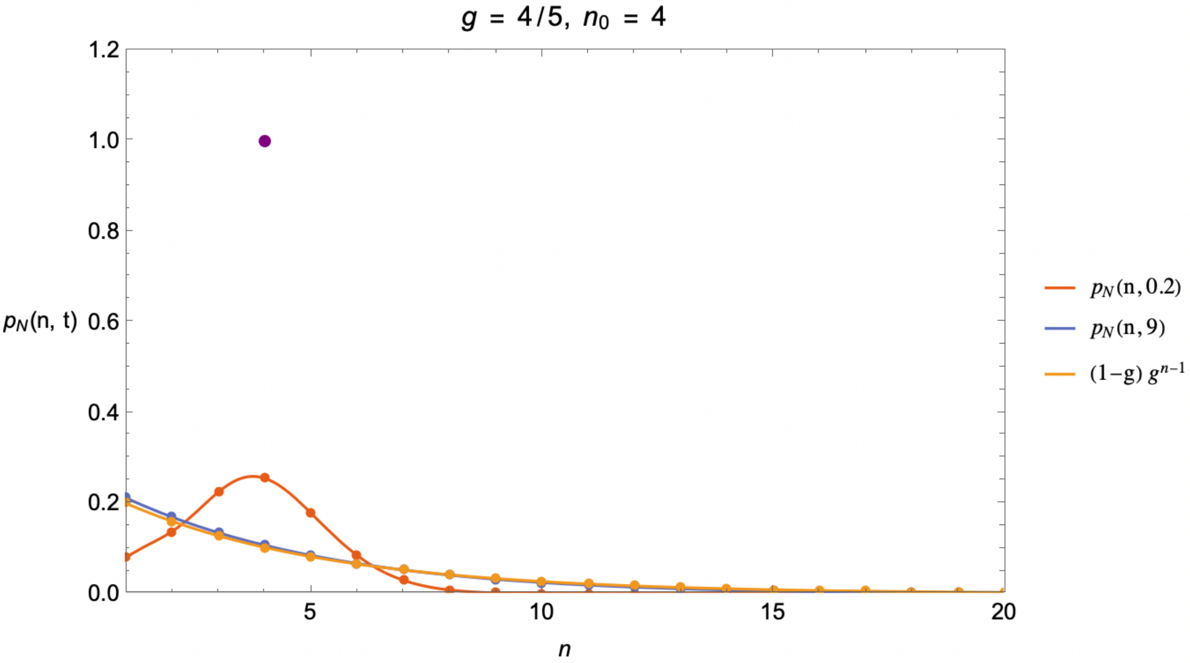

Figure 3: In this figure we represent the solutions of Eq. (26) with $\lambda = 4$ and $\mu = 5$ from which we have $\tau = 9$. Time is measured in $1/\mu$-units. The purple point represents the initial condition $p_N(n,0) = \delta_{n,4}$. The red graph represents $p_N(n,0.2)$. The blue graph represents $p_N(n,\tau)$: one observes that when $t \sim \tau$ the time dependent solution is similar to the stationary ones, represented in orange.

2 Financial markets

2.1 Introduction to financial markets

In this section we introduce some basic properties of financial markets. In particular we discuss some trading principle and we introduce the order book and some ”stylized facts”, like the distribution of asset price returns.

Financial markets are systems in which a large numbers of traders interact with one another and react to external information in order to determine the best price of a given item. These items might be as different as equities, currencies, cryptovalute, stocks, animals or others. Everything that can be traded in financial markets is generally referred to as financial asset. Some markets are localized in specific cities (New York, Tokyo, London) while others are delocalized and accessible all over the world and it is also possible to trade online.

The time evolution of the price of a financial asset is unpredictable. For example stock represents a “piece” of a company and the stock price reflects the overall value of this company. But this value depends not only on the current situation but also on the company expected future performance. If a new information is revealed that might affect the company’s future performance than the stock price will vary accordingly. This uncertainty is reflected in the erratic behaviour of the stock price that can be modeled using probabilistic tools such as stochastic processes and Brownian motion.

2.2 Orders and order book

A market is a place in which investors can buy or sell certain quantities of a financial asset. As we have seen before, the price of the asset is not static but it changes in time. The current market price, also known as market value, is defined as the price at which a share of stock last traded. The cause of the changes in the asset price are the orders investor places.

Let us consider a car. It is possible to both buy a car paying the dealer’s sticker price and to negotiate a price and refuse to finalize the deal unless the dealer meets your valuation. Stock markets work in the same way.

For an investor it is possible to make two kinds of order depending on his interests:

1. *Market orders*: they are transactions meant to execute as quickly as possible at the current market price. If a trader place a market order to buy a stock, this order will be executed immediately at the current bid price of the stock. The same will happen if he decides to buy a certain quantity of stock.

2. *Limit orders*: they set the maximum or minimum price at which the investor will to buy or sell. They are called limit orders. For example if a trader place a limit order to buy a certain quantity of the asset at a certain price, the order will be executed only when the current price of the asset will equal to this price or lower.

Because the second one needs time to be executed it is necessary to register them is a sort of register: the order book.

Order book refers to a list of buy and sell limit orders (also called respectively bid and ask) for a specific security or financial asset organized by price level. Some important features of the order book are:

- *Bid side*: the list of the buy order and it is on the left of the current price;
- *Ask side*: the list of the sell order and it is on the right of the current price;
- *Top of the book*: the highest bid and the lowest ask;

- *(Bid-ask) spread*: the difference between the top ask and the top bid. One can imagine the current price is in the middle of the spread;
- *Book/Market depth*: in this thesis we refer to this concept as the total amount of volume at each price level and it is related to the concept of liquidity, as we will see in subsection 3.2.

In Fig. 4 all these properties are graphically resumed.

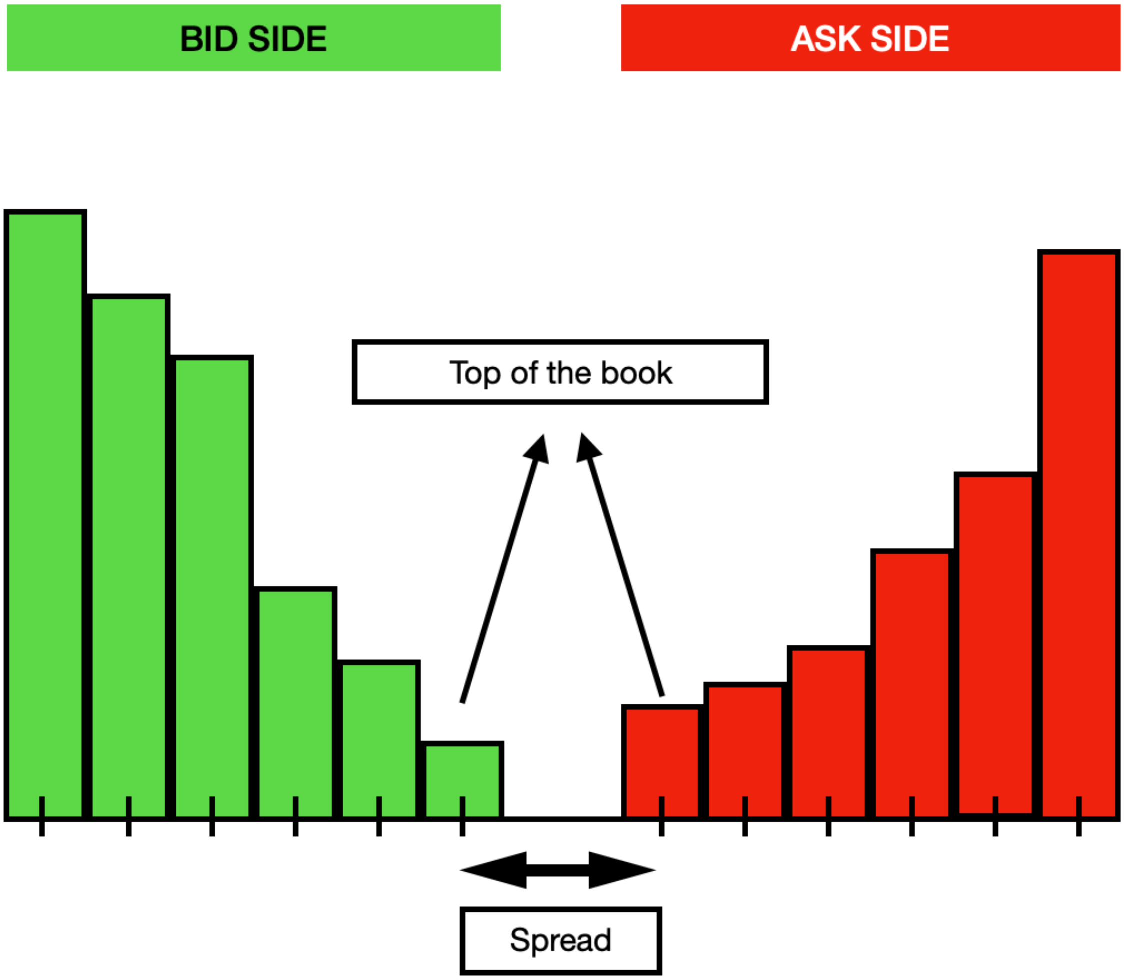

Figure 4: In this picture the cumulative depth chart of an order book is shown. The green side is the bid side: a column represents the sum of the orders at each price level from the top bid to the column itself. The red side is the ask side and the meaning of the columns is the same as before. Market/book depth can be seen as the sum of the first and the last column of the chart.

2.3 Geometric Brownian motion

Now we try to model asset prices as stochastic processes using a variant of the Brownian motion. Let $Y(t)$ be the price of a financial asset at time t and $X(t) = \ln Y(t)$ be the natural logarithm of the asset price.
The most common stochastic model of stock price dynamics assumes that $X(t)$ is a diffusive

process and the $X(t)$ increments are Gaussian distributed. This model is called "Geometric Brownian motion" [8]. Indeed, in this model, $Y(t)$ is the solution of the following stochastic differential equation:

$$dY = \omega Y dt + \sigma Y dB \, , \tag{35}$$

where ω and σ are real and positive constants. In particular σ is called volatility. Remembering the definition of $X(t)$ we can expand its differential up to the second order:

$$dX = d(\ln Y) = \frac{1}{Y} dY - \frac{1}{2Y^2} dY^2 \, . \tag{36}$$

Substituting Eq. (35) we obtain:

$$dX = \frac{1}{Y}(\omega Y dt + \sigma Y dB) - \frac{1}{2Y^2}(\omega Y dt + \sigma Y dB)(\omega Y dt + \sigma Y dB) \, . \tag{37}$$

Considering the second order differentials as negligible and knowing the Wiener process is proportional to $dt^{1/2}$ [1] we obtain:

$$dX = \omega dt + \sigma dB - \frac{\sigma^2}{2} dt = (\omega - \frac{\sigma^2}{2}) dt + \sigma dB = \omega' dt + \sigma dB \, , \tag{38}$$

where

$$\omega' = \omega - \frac{\sigma^2}{2} \, . \tag{39}$$

Defining $S(t, t_0) = X(t) - X(t_0)$ as the (log-)return of the asset price, we obtain the following expression:

$$S(t, t_0) = \omega'(t - t_0) + \sigma B(t) \, , \tag{40}$$

where $B(t_0) = 0$ and whose probability density function is [8]:

$$p_S(s, t, t_0) = \frac{1}{\sqrt{2\pi\sigma^2 \Delta t}} e^{-\frac{(s - \omega' \Delta t)^2}{2\sigma^2 \Delta t}} \, , \tag{41}$$

where $\Delta t = t - t_0$.

This model has to face empirical data, obtained from financial time series. Systematic deviations from the model predictions are observed: the distribution of returns seems to be non-Gaussian, sharp peaked and heavy tailed. This behaviour of the asset prices is one of the so called stylized facts, that we briefly describe in the following paragraph.

2.4 Stylized facts

As we have seen in the previous sections, there are different financial assets and different kind of markets. One could easily imagine that, since different markets are not necessary influenced by the same events or information sets, price series obtained from different markets will exhibit different properties. Indeed it is reasonable to think real estate, corn futures and cryptovalute markets are not similar.

Anyway, analyzing their properties from a statistical point of view one can observe that seemingly random variations of asset prices share some non trivial statistical properties. These facts, common across a wide range of assets, markets, and time periods, emerging from the statistical analysis of price variations are called stylized facts.

There's a lot of these common properties and in the following we show some of the most interesting:

- *Heavy tail*: the precise form of the tail of the distribution of returns is difficult to determine. However it seems to display a Pareto-like tail (that is, a power-law tail).
- *Aggregational Gaussianity*: increasing the time scale τ over which returns are calculated their distribution appear more like a Gaussian distribution.
- *Volume/volatility correlation*: trading volume, that is the number of shares, is correlated with all measures of volatility.

There are other stylized facts such as the asymmetry of the return distribution and the absence of autocorrelations. See for instance [7].

2.5 Returns distribution properties and models

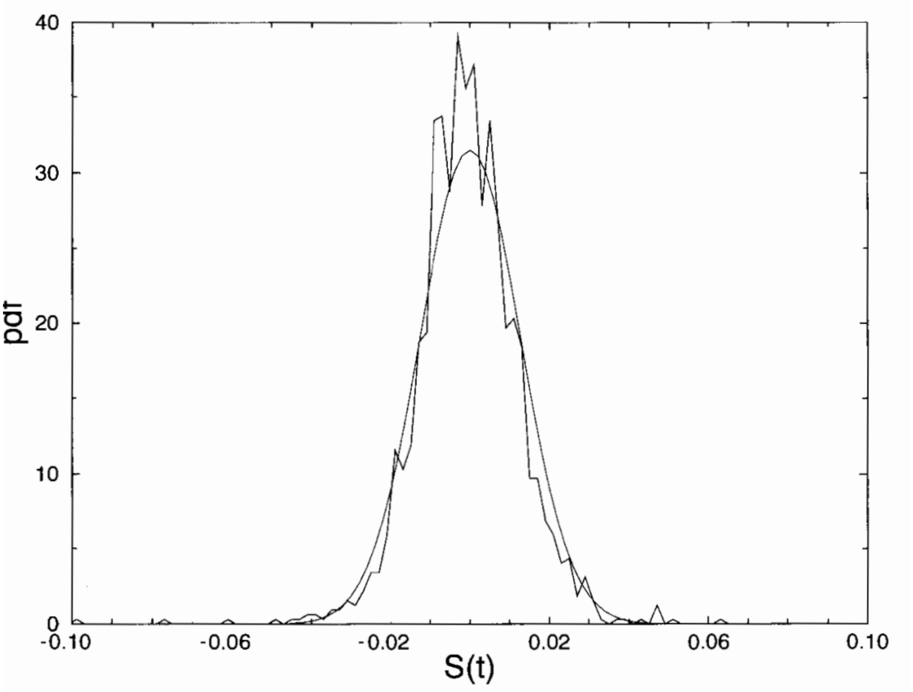

Figure 5: Empirical PDF for log-returns of Chevron stock traded in New York Stock Exchange in the period 1989 to 1995. The smooth line is the Gaussian distribution with the same variance calculated from the data [6].

We want to study in more detail the properties of the asset return distribution. In Fig. 5 we show an example of this kind of distribution.

One way to quantify the deviation from the normal distribution is by using the coefficient of kurtosis κ defined as the ratio of the fourth central moment and the square of the second central moment [4]:

$$\kappa = \frac{\mathbb{E}[(S(t,t_0) - \mathbb{E}[S(t,t_0)])^4]}{\mathbb{E}[(S(t,t_0) - \mathbb{E}[S(t,t_0)])^2]^2} \,. \tag{42}$$

- If $\kappa = 3$ we have a Gaussian distribution;
- If $\kappa > 3$ we have a "fat tails" distribution and it is said leptokurtic;
- If $\kappa < 3$ the distribution is said platikurtic.

Analyzing different sets of returns, at different time scales, one can observe that κ spread from 15 to higher values such as 60 or 74 and that these properties are more pronounced if the time scale $\Delta t < 1$ day that is for intraday values of Δt [7]. There have been proposed many parametric models trying to explain the return distribution.

The Levy stable non-Gaussian model implies that $\ln Y(t)$ undergoes a discontinuous time evolution and $S(t, t_0)$ is characterized by a non-Gaussian distribution with infinite second and higher moments [6][8].

Another model that has been proposed uses the Student's t-distribution. This distribution is governed by a parameter n and the k-moments of the distribution with $k < n$ are all finite. By varying the control parameter n one can approximate with good accuracy the log price change distribution [6].

A different model proposed by Clark utilized the concepts of mixture of Gaussian distributions and subordination between stochastic processes. Clark interpreted the leptokurtic behaviour observed in empirical analysis as the result of the fact that the trading activity is not uniform during the trading interval. In fact from analysis of the market activity it is known that the number of transactions occurring in the market in a given time fluctuates. We call this fluctuating number trading volume and we denote the cumulative trading volume up to time t by $\Omega(t)$. In his model the distribution of log price increments occurring from a given level of trading volume $P(S(\Omega(t))$ is subordinated to the one of individual trade $P(S(t))$ and directed by the distribution of the trading volume. By assuming $P(S(t))$ to be Gaussian and $P(\Omega)$ to have all moments finite, Clark proved that $P(S(\Omega(t)))$ is a leptokurtic distribution with all the moments finite. The specific form of the distribution depends on $\Omega(t)$ [6].

The last model is interesting because it tries to connect the distribution of returns to market microstructure properties such as trading activity. In the next section we will show a simple model following the idea that there exists a process subordinated by another "microscopic process".

3 Polymer Physics, financial markets and anomalous diffusion

3.1 Analogies between polymer physics and financial markets

In this section we show the principal result of this thesis work: the proposal of a model based on queueing theory for the order book in order to describe the time variations of asset prices.
It is possible to make a parallel between a polymer and an order book:

- Both are of course two chains or queues, whose beads are called respectively monomers and limit orders;
- The polymer center of mass is related to the current price that can be thought in the middle of the spread;
- The size of the polymer is related to the market depth;
- Brownian motion, which every monomer is subjected, can be related to the geometric Brownian motion for the returns;
- A chemostatted polymer, that is a polymer immersed in a monomers bath, can be related to an orders book immersed in an "order bath";
- Polymerization process, the association of a monomer to the chain, is equivalent to the arrival of a new limit order;
- Depolymerization process, the departure of a monomer from the chain, is equivalent to the execution of a limit order.

In several models (like Rouse, Zimm and Reptation) the volatility of the polymer center of mass is related to the polymer size (see Eq. (21)). In order to complete the analogy between these two systems, it should be natural to wonder if the volatility of an asset price is somehow related to the book/market depth.

3.2 Volatility and liquidity

Market liquidity refers to the efficiency or ease with which an asset or security can be converted into ready cash without affecting its market price. In other words, the liquidity of the market can be recognised by how low the uncertainties of execution price are. The most liquid asset is cash itself; real estate market is an example of illiquid market.

In addition, one consider market depth, which absorbs the price changes accompanied by trade execution, as an important factor in explaining market liquidity. Then, market depth can be thought as a liquidity proxy.

Now, it is well known [9] in market microstructure theory that liquidity (market depth) and assets volatility are inverse correlated or anticorrelated. This means the deepest the market, the lower the volatility and vice versa. It is reasonable to model the relationship between volatility of the asset price and market depth as:

$$\sigma = \sqrt{2D(N)} = \sqrt{2D_0/N^{\alpha}}\,, \tag{43}$$

where N is the total number of limit orders and $\alpha > 0$ is a real number (in analogy with Eq. 22).

3.3 A simple birth and death model for the order book

In a critical chemostatted polymer, size is time dependent because of polymerization, and this is one of the two sources of randomness (the other one is the standard thermal agitation imparted by the solvent and represented by the Wiener process). The size of the polymer can be described with a birth-death process: we try to do the same with the "size" of the order book (the book depth).

If we think the orders listed in the order book as a population we can easily apply birth and death processes to his description [10][11]. A simple birth and death model for the order book can be built with these assumptions:

- All orders are assumed to be of unit size;
- Price levels are discrete, $\{1, 2, \ldots, k\}$;
- We consider only a side of the book, for example the ask side. So all limit orders are ask orders and all market orders are buy orders;
- For all $p \in \{1, 2, \ldots, k\}$ limit orders at price p are submitted according to a Poisson process with parameter λ_p . These processes are assumed to be mutually independent so that the number of orders submitted is a Poisson process with $\lambda = \sum_p \lambda_p$;
- Limit orders are executed according to a Poisson process with parameter μ.

In this context, $N(t)$ represents the book depth, the number of orders listed in the order book at time t. The state probabilities follow Eq. (26) with size independent birth and death rate, that is the same equation governing the polymerization/depolymerization process in the mean field limit.

3.4 Anomalous diffusion

Since we are considering two systems sharing the same phenomenology, we denote both the polymer center of mass and the logarithm of the asset price with X. The size of the system is indicated by N.

Since $N(t)$ is a random variable, also the coefficient of diffusion has to be (see Eqs. (21) and (43)). This means that in the evolution of the variable $X(t)$ there are two source of randomness:

1. The thermal agitation imparted in the first situation by the fluid in which Brownian particle is immersed and in the second by the interactions between traders. It is represented respectively by Brownian motion and geometric Brownian motion.
2. The polymerization/depolymerization process which affect the intensity of the Brownian motion and the arrival/departure of new limit orders which affect the intensity of the geometric Brownian motion.

Brownian motion is named the subordinated process and $N(t)$ the subordinator process.

Remembering Eqs. (22) and (38) (with $\omega' = 0$), X evolves according to:

$$dX(t) = \sqrt{2D(N)}dB(dt)\,. \tag{44}$$

Under ordinary conditions the stationary distribution of N is strongly peaked around its mean value and the "diffusion of the diffusive coefficient" is difficult to detect. But in proximity of the critical point (that is, when $g \to 1^-$) the divergence of the size of the system N changes drastically the situation.

In the limit $t \to 0^+$ it is possible to study the subordinator character of the diffusion. For any given realization $[n(t)] = \{n(t') \in \mathbb{N}^* | 0 \leq t' \leq t\}$ of the stochastic process $N(t)$ the PDF of X satisfies this diffusion equation:

$$\frac{\partial}{\partial t} p_X(x, t|[n(t)]) = \frac{D_0}{(n(t))^\alpha} \frac{\partial^2}{\partial x^2} p_X(x, t|[n(t)]) \,, \tag{45}$$

with the initial condition $X(0) = 0$. Now we reparameterize the diffusing path in terms of a coordinate which converts the diffusion coefficient in an unitary one:

$$dX(z) = dB(dz) \,, \qquad dz = 2D(n(t))dt \,, \tag{46}$$

where $z \geq 0$ is a path variable corresponding to the realization of the stochastic process $Z(t)$ defined as:

$$Z(t) = \int_0^t dt' 2D(N(t')) \,. \tag{47}$$

Now we can rewrite Eq. (45) as:

$$\frac{\partial}{\partial z} p_X(x, z) = \frac{\partial^2}{\partial x^2} p_X(x, z) \,, \tag{48}$$

whose solution is:

$$p_X(x, z) = \frac{e^{-\frac{x^2}{2z}}}{(2\pi z)^{1/2}} \,. \tag{49}$$

In order to determine the probability $p_X(x, t)$ of finding the center of mass at position x at time t, knowing it starts in the origin of the frame of reference with $N(0) = n_0$ monomers, we can use the subordination formula:

$$p_X(x, t) = \int_0^\infty dz \frac{e^{-\frac{x^2}{2z}}}{(2\pi z)^{1/2}} p_Z(z, t) \,, \tag{50}$$

where $p_Z(z, t)$ is the probability distribution of the process $Z(t)$. This formula describes $p_X(x, t)$ as a mixture of Gaussian distributions.

At large time $t \gg \tau$ the subordinated Brownian character becomes distinctive: since D is deterministic, then in this limit $Z(t)/t$ is a number itself [4]. One can think the diffusion coefficient as the following weighted mean:

$$D_{av} = \sum_{n=1}^{\infty} D(n) p_N^*(n) = \sum_{n=1}^{\infty} \frac{D_0}{n^\alpha} p_N^*(n) \,. \tag{51}$$

In particular it is interesting to look at the shape of the initial probability density function of X under equilibrium conditions. If one defines the unit-variance variable:

$$\overline{X}(t) = \frac{X}{\sqrt{\mathbb{E}[X^2(t)]}} \,, \tag{52}$$

it can be shown (see App. A), using subordination formula and after some algebraic calculations involving path integrals, that in the limit $t \to 0^+$ $\overline{X}(t)$ follows the PDF:

$$p_{\overline{X}}(\overline{x}, 0^+) = \sum_{n=1}^{\infty} p_N^*(n) \frac{e^{-\frac{D_{av} n^\alpha \overline{x}^2}{2D_0}}}{\sqrt{2\pi \frac{D_0}{D_{av} n^\alpha}}} \,, \tag{53}$$

which depends only on g and α. In Fig. 6 we show the graph of this distribution.

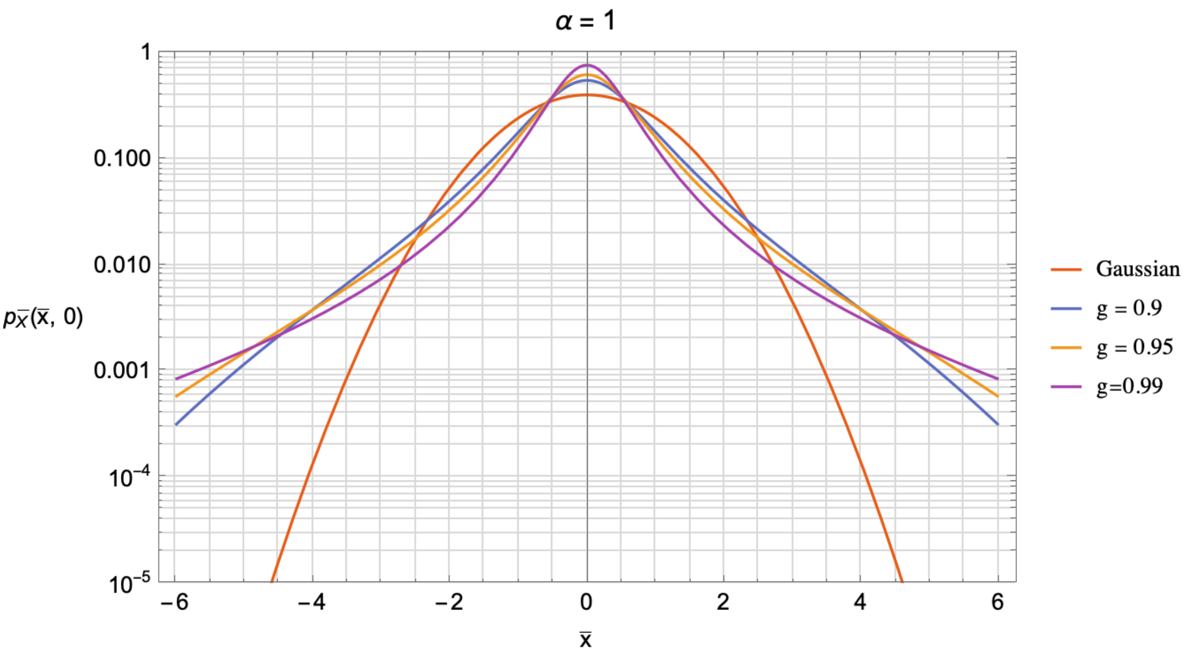

Figure 6: In this picture we show the semi-log graph of the distribution Eq. (53) for different values of g when $\alpha = 1$, that is for a Rouse polymer. Also the unit-variance Gaussian is plotted.

As we can see from the graph in Fig. 6, the distribution $p_{\overline{X}}(\overline{x}, 0^+)$ is sharp peaked, heavy tailed and leptokurtic when $g \to 1^-$. For all the α coefficients the kurtosis diverges at the critical value $g = 1$ at time $t = 0^+$. Comparing Figs. 6 and 7 one can appreciate the similarity between the distribution predicted by this model and an empirical distribution of log returns.
Instead, when $t \gg \tau$ it can be shown [4] that a power law decay holds for the excess kurtosis: $\kappa - 3 \propto 1/t$. This means as the time increases, $p_{\overline{X}}(\overline{x}, 0^+)$ appears more Gaussian, as expected from empirical studies on the distribution of log-returns.

3.5 Conclusions

In this thesis we have briefly introduced some basic concepts of Polymer Physics and financial markets. In the last section we have shown that these two worlds seem to share the same phenomenology. This allowed us to derive the PDF of both the center of mass of a chemostatted polymer near the critical polymer in the mean field limit and the log-returns of a financial asset. In particular, the two key points are the hypothesis on the relation between volatility and liquidity and the birth-death model of the order book. It would be very interesting to study in more detail how market depth affects asset volatility, for instance including imbalances between bids and asks. Other models could be proposed using more complex queues than the M/M/1 or trying to model both sides of the order book.
Now the model we have proposed has to face empirical data: the next step is the calibration of the model through financial data. Thanks to this procedure it will be possible to verify its goodness and to individuate the best parameters g and D_0 in order to fit empirical data.

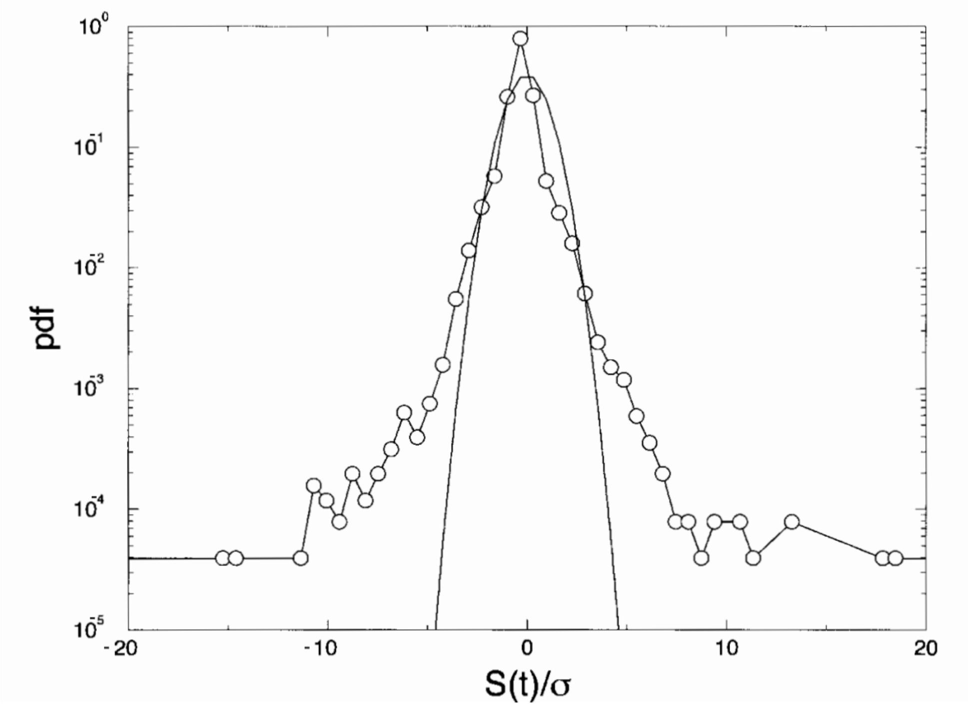

Figure 7: Empirical PDF for high frequency log-returns of the Xerox stock traded in the New York Stock Exchange during the two years period 1994 to 1995. The smooth line is the Gaussian distribution with the same variance calculated from the data. The plot is semi-logarithmic [6].

A Demonstration of Eq. (53)

Since the stochastic process $Z(t)$ (defined in Eq. (47)) depends on $N(t)$ we can express the PDF $p_Z(z,t)$ of the former in term of the PDF $p_N[n(t)]$ of the latter:

$$p_Z(z,t|n_0) = \int d[n(t)]\delta\left(z - \int_0^t dt' \frac{2D_0}{(n(t'))^\alpha}\right) p_N[n(t)]\,, \tag{54}$$

with initial condition $N(0) = n_0$. Then, from Eq. (50),

$$\begin{aligned} p_X(x,t|n_0) &= \int d[n(t)] \int_0^\infty dz \frac{e^{-\frac{x^2}{2z}}}{(2\pi z)^{1/2}} \delta\left(z - \int_0^t dt' \frac{2D_0}{(n(t'))^\alpha}\right) p_N[n(t)] = \\ &= \int d[n(t)] \frac{e^{-\frac{x^2}{2\int_0^t dt' \frac{2D_0}{(n(t'))^\alpha}}}}{\left(2\pi \int_0^t dt' \frac{2D_0}{(n(t'))^\alpha}\right)^{1/2}} p_N[n(t)]\,. \end{aligned} \tag{55}$$

By discretizing the time in k_m intervals lasting Δt and using the Markovianity of the process $N(t)$ we can take the limit $k_m \to \infty$, $\Delta t \to 0$ such that $k_m \Delta t = t$:

$$p_X(x,t|n_0) = \lim_{k_m\to\infty} \sum_{n_1,\ldots,n_{k_m}} p_N(n_{k_m}|n_{k_{m-1}}) \ldots p_N(n_1|n_0) \frac{e^{-\frac{x^2}{2\sum_{k=1}^{k_m} \Delta t \frac{2D_0}{(n_k)^\alpha}}}}{\left(2\pi \sum_{k=1}^{k_m} \Delta t \frac{2D_0}{(n_k)^\alpha}\right)^{1/2}}\,. \tag{56}$$

Since we are interested in the limit $t \to 0^+$ we consider only the first time interval Δt:

$$p_X(x,\Delta t|n_0) = \sum_{n_1=1}^\infty p_N(n_1|n_0) \frac{e^{-\frac{x^2}{2\Delta t \frac{2D_0}{(n_k)^\alpha}}}}{\left(2\pi \Delta t \frac{2D_0}{(n_k)^\alpha}\right)^{1/2}}\,. \tag{57}$$

Eliminating the dependence from the initial condition:

$$p_X(x,\Delta t) = \sum_{n_0=1}^\infty p_X(x,\Delta t|n_0) p_N^*(n_0). \tag{58}$$

If we now use Eq. (57):

$$p_X(x,\Delta t) = \sum_{n_1=1}^\infty p_N^*(n_1) \frac{e^{-\frac{x^2}{2\Delta t \frac{2D_0}{(n_1)^\alpha}}}}{\left(2\pi \Delta t \frac{2D_0}{(n_1)^\alpha}\right)^{1/2}}\,. \tag{59}$$

From the previous it follows that:

$$\mathbb{E}[X^2] = \sum_{n_1=1}^\infty p_N^*(n_1) \frac{2D_0 \Delta t}{(n_1)^\alpha} = 2D_{av}\Delta t\,, \tag{60}$$

where D_{av} is defined as in Eq. (51). Now passing to the variable $\overline{X} = X/\sqrt{\mathbb{E}[X^2]}$ we obtain:

$$p_{\overline{X}}(\overline{x},\Delta t) = \int dx p_X(x,\Delta t)\delta\left(\overline{x} - \frac{x}{\sqrt{\mathbb{E}[X^2]}}\right) = \int dx p_X(x,\Delta t)\delta\left(\overline{x} - \frac{x}{\sqrt{2D_{av}\Delta t}}\right)\,. \tag{61}$$

From the latter we finally obtain Eq. (53).

B References

[1] Daniel T. Gillespie, 1996. The mathematics of Brownian motion and Johnson noise. *American Journal of Physics* 64, 225.

[2] W.J. Briels, October 1998. *Theory of Polymer Dynamics.*

[3] C. Vanderzande, 1998. *Lattice Models of Polymers.* Limburgs Universitair Centrum, Belgium.

[4] Sankaran Nampoothiri et al 2022 *New J. Phys.* 24 023003.

[5] Artem S. Novozhilov, Georgy P. Karev and Eugene V. Koonin, 2005. Biological applications of the theory of birth-and-death processes. *Briefings in bioinformatics.* Vol 7. No I. 70 - 85.

[6] Rosario N. Mantegna and H. Eugene Stanley, 2000. *An introduction to Econophysics. Correlation and Complexity in Finance.* Cambridge Univeritiy Press.

[7] Rama Cont, 2000. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, Volume 1 (2001) 223 - 236.

[8] Giovani L. Vasconcelos, 2004. A Guided Walk Down Wall Street: An Introduction to Econophysics. *Brazilian Journal of Physics*, vol. 34, no. 3B, September, 2004.

[9] Namitha K. Cheriyan, Daniel Lazar, 2018. Relationship between Liquidity, Volatility and Trading Activity: An Intraday Analysis of Indian Stock Market. *International Journal of Economics and Financial Issues*, 2019, 9(1), 17-22.

[10] Ioane Muni Toke. The order book as a queueing system: average depth and influence of the size of limit orders. *Quantitative Finance, Taylor and Francis* (Routledge), 2015, 15 (5), pp.795-808.

[11] Cont, Rama, De Larrard, Adrien. 2011. Price dynamics in a Markovian limit order market. *arXiv:1104.4596*, Apr.