Universidad del País Vasco
Euskal Herriko Unibertsitatea
ZIENTZIA ETA TEKNOLOGIA FAKULTATEA
FACULTAD DE CIENCIA Y TECNOLOGÍA

Gradu amaierako lana /Trabajo fin de grado
Fisikako gradua/ Grado en Física

Introduction to Fundamental Techniques in Econophysics

with Python Simulations

Egilea/ Autor/a:
Jon Rodriguez Gonzalez
Zuzendariak/Directores/as:
Josu Mirena Igartua

Leioa, 2025eko ekainaren 18a / Leioa, 18 de junio de 2025

Table of contents

1 Introduction

In this work we will try to deep into the basics of econophysics. For that we shall start by introducing a key concept in economic markets: stochasticity. We will see its relation to physical systems and introduce basic physical equations which can be used to describe this type of markets.

We will see that the first approach to describing financial markets by physical systems leads to the use of Brownian motion. However, its results adhere to the dynamics of financial markets only under very specific conditions. Crucially, it fails to accurately describe sudden price movements or extreme events, which are of paramount importance in understanding financial markets. In this paper, we aim to address these limitations by exploring alternative stochastic models.

We will see that although better suited alternatives appear, such as Levy flights, this type of models lack various properties, mainly the description of single markets without taking into consideration the relationship amongst all stocks.

In this regard we will redirect our problem to obtaining models that consider the intricate relationships between financial markets. For that we shall firstly see that this correlation [9] indeed appears by means of a correlation matrix.

We will also make use of the analogy with nuclear systems, as for we consider both those and financial systems complex systems which can not be described trough the single interactions between all agents taking place, but rather more phenomenological theories need to be developed.

Finally a natural transition towards modern techniques for describing markets will be made, mainly: Random Matrix Theory and Portfolio Theory.

The aim of this work, thus, is to see why basic notions such as Brownian or Levy models are to be discarded when aiming to correctly describe markets and to address the necessity to obtain models that take into account the interactions amongst markets. It is important to highlight that this work does not contain any self-developed theories nor is the intention to proclaim any of them, but rather to use information from the literature as well as self-developed python codes to support main arguments and create a unique structure to explain the basics of econophysics. All the necessary references will be made all along to address the reader to the raised topics.

There will be no clear theoretical introduction, as the needed concepts will be explained in further detail in each section. In case further theoretical background is needed by the reader, we address the references provided at the end of this document. These references include foundational works on stochastic processes, econophysics, and financial modeling, which will help the reader gain a deeper understanding of the topics discussed.

Note: The extended work and all Python scripts are available in the GitHub repository under the name "GrAL".

2 Motivation

Physicists have long been interested in solving hidden complex structures in nature. From basic kinematic equations to the intricacies of quantum and statistical systems, there is always an urge to understand how the world functions and give a clear description of its behavior. That is so, fields that would not clearly be linked to physical studies have been embraced by physicists, who have taken on the task of modeling them. Some of them include the now very popular artificial intelligence and data science, where statistical laws are being applied to obtain precise models. Notably, physicists are embarking on other new areas such as finance; what used to be an area of study of social sciences is now starting to become more scientifically modeled by rigorous procedures.

It is in this realm that we try to give an introduction to the state-of-the-art of the commonly known "econophysics". This work is a natural evolution of the work carried out by Naia Ormaza Zulueta in her Final Degree Project (TFG) [15], where we expand on the concepts already introduced there and give a logical progression towards the status of this field nowadays. We will see that most of the arguments presented by Naia are to be re-evaluated and refined to arrive at new more accurate methods to describe financial systems.

Note that the aim of this work is not to find innovative procedures that could be applied in this field, not the least because its complexity would require a far deeper analysis of the topic. Experts across the world are devoting their time to these challenges. Thus, the focus of this paper will be to try and transition from past old fashioned financial models towards innovative investigations, guiding the process and explaining the foundations to understand this new emerging and exhilarating discipline.

On a personal note, I feel deeply captivated by the interdisciplinary avenues that open up when physics is applied thoughtfully. I believe there is always a way to interpret even the most complex phenomena, and I'm excited by the challenge of uncovering those connections to bring clarity and insight to systems that at first glance seem inscrutable.

3 The importance of stochastic models

Stochastic processes are mathematical concepts used to describe random variables that change over time, so their main attribute is that of being random. They are widely used in physics, biology, economics, and other fields to model complex systems that are difficult to predict. In particular, various physical systems are developed from stochastic principles. It is the aim of this chapter to introduce some of them, to give account of the importance of stochasticity in the study of real-world systems. We will start by showing that both classical and quantum noise are stochastic processes and that their properties are developed from this feature [5]. Then we will see that the motion of particles in a fluid (mainly, the Brownian motion) also follows stochastic patterns. Furthermore, this model will be of great use when studying financial markets, so it serves as a good introduction to the topic

3.1 Classical and quantum noise

Quantum noise is gaining importance as its influence on quantum computers is becoming more noticeable, disturbing the states of qubits. Firstly, though, what are qubits? Qubits are the analogues of bits in regular computers, although they differ mainly in the amount of states or positions they can be found in. It is already common knowledge that bits can take values of 0 or 1, depending on the charge of the capacitors they associate with: one "charged" capacitor leads to a binary 1 while a discharged capacitor means a binary 0. Similarly qubits can be found to be in two states, which we ordinarily denote according to the quantum Dirac description $|0\rangle$ and $|1\rangle$, as well as we have seen with the ordinary 0 or 1 for bits. However, contrary to the classical description of bits, qubits obey the laws of quantum mechanics and can therefore also be described as a superposition of their base quantum states that we denote by $|\psi\rangle = a|0\rangle + b|1\rangle$. As the total probability needs to be 1 we normalize the result by imposing that the sum of the squared modulus of a and b be 1.

Quantum noise, then, disturbs the states of qubits and produces an effect known as quantum decoherence where a quantum entangled state becomes a classical not entangled state, where quantum effects are no longer shown and classical behavior is thus observed. It is the stochastic process which produces the commonly known shot noise, which differs from the Johnson-Nyquist noise found in equilibrium currents. This noise appears as a result of the discrete behavior of energy carriers which leads to fluctuations in the measurements. It has an important effect both in electronic circuits, where the energy carriers are electrons, and in optical applications, where photons are the ones responsible for carrying energy.

3.1.1 Classical noise

Noise is key in obtaining the desired results and it can be a significant challenge to deal with, so it is important that we quantify its effects. We start by developing the classical noise formulas and then obtaining its quantum versions. For that let us consider a Gaussian-like distribution for classical noise which we denote by $I(t)$, following the central limit theorem. We will also accept that the autocorrelation vanishes when a time

period τ has elapsed and that statistically there are no changes when time is shifted, that is, that stationarity [9] is assumed. We can define the autocorrelation function to be

$$G(t, t_0) = \langle I(t)I(t_0)\rangle = G(\tau) = \langle I(t)I(t+\tau)\rangle \tag{1}$$

The spectral density, which we will denote by $S[w]$, is attained by taking the limit of the power of the Fourier transform as the period approaches infinity, a measure which shows us how the energy or power of the intensity is distributed across all the frequencies that appear in the signal [18]. From the Wiener-Khinchin theorem we conclude that the Fourier transform of $G(t)$ also leads to to the same results. Thus, in account of the properties of $G(t)$ we get that $S[w] = S[-w]$, of course, symmetric in frequency. This is a very important conclusion for classical noise and obtained from the premise that it is stochastic.

To see that the symmetry described above for classical noise systems truly exist, let us simulate our own classical system . For that we will consider a fairly simple electronic circuit composed of one resistor and one capacitor in series and a voltage source which will be the input of the thermal random noise.

This is one of the most common electronic circuits, being a linear time invariant circuit widely used to filter AC signals. The voltage across the capacitor is given by:

$$V_C(t) = \int \frac{I(t)}{C} dt \tag{2}$$

and in taking the spectral density of the current, we obtain, for $t = t'$:

$$\langle I(t)I(t')\rangle = \frac{2k_B T}{R} \tag{3}$$

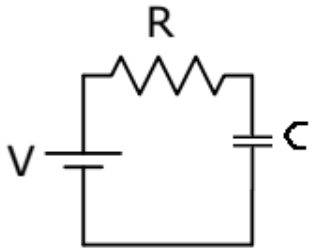

Figure 1: RC circuit where the input voltage is a random thermal noise

Note that the spectral density shows how the energy is distributed in the frequency spectrum and if linearity is assumed then we do not observe phase changes. Therefore we can obtain the system‘s spectral density [5] by

$$S_{output}(\omega) = S_{input}(\omega)H(\omega)^2 = S_{\mathrm{RC}}(\omega) = \frac{2k_B TR}{1 + (\omega RC)^2} \tag{4}$$

so that the output spectral density is the input spectral density multiplied by the square of the transfer function, $H(w)$. This is a very important result as it shows that the output of the system is directly related to the input and the transfer function, which is a key concept in signal processing. The graph of this function is shown below:

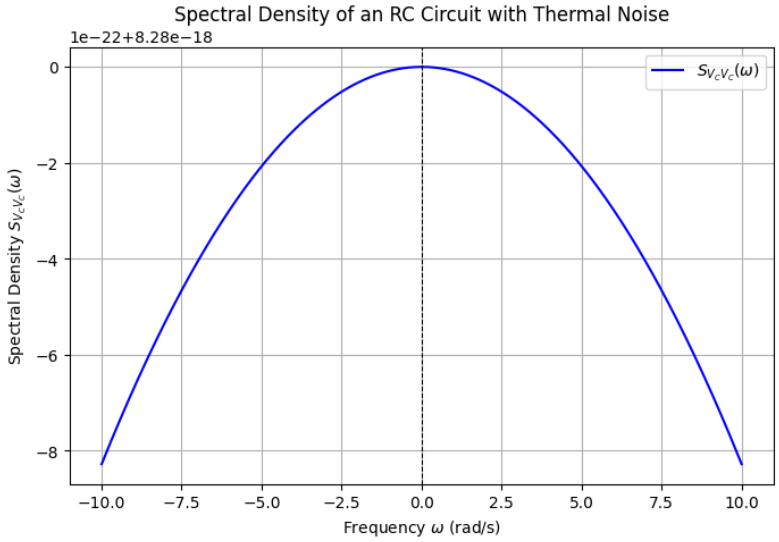

Figure 2: Spectral density of S_{RC} for an RC circuit with thermal noise, showing a Lorentzian profile.

The physical system described above where the input is a stochastic noise source is not the only one to show this symmetry. It can also be seen easily by considering one of the most known systems: the harmonic oscillator in thermal equilibrium. Its autocorrelation function is given by

$$G_{xx} = \frac{k_B T}{M\Omega^2} \cos(\Omega t) \tag{5}$$

which leads to [5] a spectral density :

$$S_{xx}[\omega] = \pi \frac{k_B T}{M\Omega^2} \left[\delta(\omega - \Omega) + \delta(\omega + \Omega)\right] \tag{6}$$

The symmetry around the y axis is clearly visible when simulating this motion:

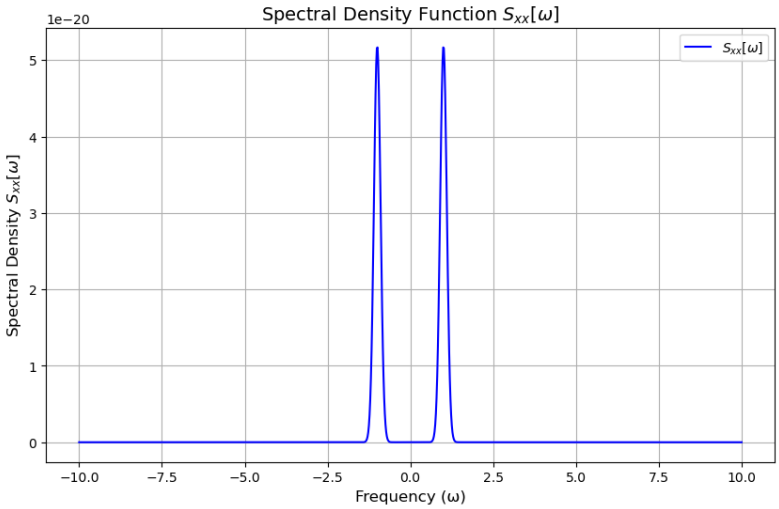

Figure 3: Spectral density function $S_{xx}[\omega]$ for a harmonic oscillator in thermal equilibrium where M=1 and T=300 K. The frequency is set to 1rad/s.

We clearly see, then, that for systems with input being classical stochastic noise we obtain certain common properties, which in our case are symmetric spectral densities.

3.1.2 Quantum noise

To transition from classical noise to quantum noise we need to introduce the Hermitian operator $\hat{I}^\dagger = \hat{I}$. Recall that an operator,$\hat{O}$, is said to be Hermitian if it holds that $\langle\psi|\hat{O}\phi\rangle = \langle\hat{O}\psi|\phi\rangle$ for all states, which guarantees that all eigenvalues are real and the eigenvectors corresponding to different eigenvalues are orthogonal. In analogy to the classical noise, its quantum spectral density is thus given by [5] :

$$S_{II}[\omega] = \int_{-\infty}^{\infty} \langle\hat{I}(t)\hat{I}(0)\rangle e^{i\omega t} dt \tag{7}$$

This differs from its classical counterpart in several ways. Firstly, when temperature is set to 0 K we observe that in the classical model the spectral density is 0, which indicates that there is no noise in the system. Conversely, the quantum model counterpart of the spectral density shows that the result is certainly not 0, a consequence of the Heisenberg uncertainty principle.

Additionally, in the quantum scheme $\hat{I}(t_1)$ and $\hat{I}(t_2)$ may not commute, that is, $[\hat{I}(t_1), \hat{I}(t_2)] \neq 0$. In contrast, when two operators do commute, their quantities can be known precisely; however, if they do not commute, their quantities cannot be simultaneously known with arbitrary precision but follow the Heisenberg uncertainty principle.

A straightforward conclusion is that the quantum autocorrelation function can be complex, and thus symmetry in frequency is not required, as it is in the classical model, a key difference between the two models. Lastly, due to the quantum nature and the Heisenberg uncertainty principle, there are inherent limits on the precision with which a system (such as an amplifier or detector) can measure signals while maintaining quantum consistency. This effect is absent in classical systems.

The same system studied for the classical regime (harmonic oscillator in thermal equilibrium) will be brought back to analyze the spectral density function‘s symmetry and we will see what the quantum results look like. We can yield the Fourier Transform of the autocorrelation function to obtain the spectral density [5]:

$$S_{xx}(\omega) = \left(\frac{\pi\hbar}{M\Omega}\right)[n_{BE}(\hbar\Omega)\delta(\omega-\Omega) + [n_{BE}(\hbar\Omega)+1]\delta(\omega+\Omega)] \tag{8}$$

At low temperatures it only remains the term $\delta(\omega+\Omega)$, as its negative counterpart vanishes due to the Bose-Einstein relation (see that for low T n_{BE} approximates to 0). This results leads to an asymmetry being introduced in the spectral density function as T falls to 0, in accordance with the theoretical conclusions we obtained earlier. It is only when temperatures increase so that the relation $k_B T \gg \hbar\Omega$ is satisfied that we cannot neglect the term n_{BE} and so the spectral density becomes symmetric, exactly as we obtained in the classical model and in good agreement with the theoretical results.

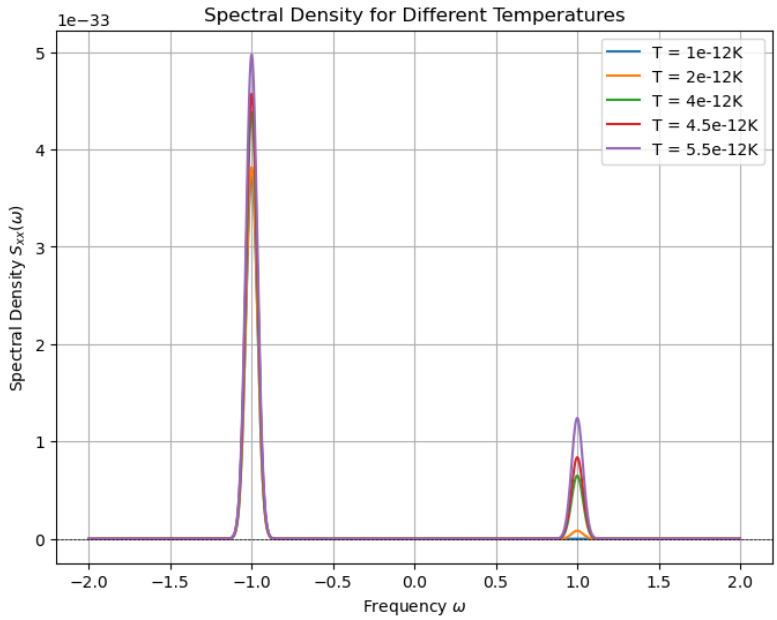

Figure 4: Spectral density function $S_{xx}[\omega]$ for a quantum harmonic oscillator in thermal equilibrium where M=1. The frequency is set to 1rad/s. The plot shows results for different very low temperatures (e.g., T values from $10^{-12}K$ to $5.5 \cdot 10^{-12}K$, as indicated in the legend), illustrating the asymmetry at these quantum regime temperatures.

At temperature $T = 10^{-12}$ the graph is antisymmetric in the y axis, but as temperatures increase, the Dirac delta starts growing for $-\Omega$, showing that we start losing symmetry very rapidly from $T = 10^{-12}$ onwards. This is in accordance with theoretical results which say that the transition should happen for values of $T \sim \frac{\hbar\Omega}{k_B}$. So for $\Omega = 1$ we get $T = 7,63823 \cdot 10^{-12}$, which is the same value we computed experimentally.

Therefore, the key takeaway point of this noise analysis is that, although their properties are different, both classical and quantum noise are stochastic processes and so their physical descriptions are obtained taking into account this stochastic behavior. Also, in sight of this randomness, they can be described by the same mathematical tools developed for the study of other stochastic models such as correlation functions, autocorrelation, and spectral density functions. This is of paramount importance as it allows us to use the same tools to study both classical and quantum systems, and thus to compare the results obtained for both models.

3.2 Introduction to Brownian motion

In this section we will introduce Brownian motion, which is also the first basic approach used for describing financial markets. It is originally used to describe the random movement of particles in a medium and is considered appropriate for financial markets because these markets are also thought to follow random patterns.

The Brownian motion is nothing more than the continuous limit of a random walk [17]. A random walk with i.i.d. steps of mean zero and unit variance, when properly rescaled in both time and step size using the factor $\frac{1}{\sqrt{n}}$, converges to Brownian motion as $n \to \infty$. Donsker's Theorem rigorously establishes that the finite-dimensional distribu-

tions and the overall path behavior of the random walk converge to those of the Wiener process [7]. Thus, Brownian motion can be understood as the continuous limit of a discrete random walk, a result firmly grounded in the Central Limit Theorem and its functional extension provided by Donsker's Theorem. The validity of these arguments are further developed in the extensive version of this work, should the reader find interesting its mathematical reasoning.

Let us now try to deduce the Brownian dynamic equations and compare them to its analytical solutions. The motion of particles in a medium under short timescales can be approximated by the diffusion equation, representing the likelihood of finding a particle at a given position in time. This particle motion is caused as a result of the collision amongst all constituents of the medium which causes their paths to be unpredictable. Indeed, it is this erratic movement that we define as Brownian motion. The diffusion equation writes as :

$$\frac{\partial P(\mathbf{r},t)}{\partial t} = \nabla \cdot \left[D(P,\mathbf{r}) \nabla P(\mathbf{r},t) \right], \tag{9}$$

where P(r, t) is the probability density of finding the particle at position x at time t and D(P, r) is the collective diffusion coefficient. It is derived from the continuity equation and it describes how the probability density function evolves over time. Its solution is given by a Gaussian distribution, which accounting for initial conditions, is given (in 1 dimension) by:

$$P(x,t) = \frac{1}{\sqrt{4\pi D t}} \exp\left(-\frac{x^2}{4Dt}\right) \tag{10}$$

This argument holds in the short timescale approximation. Yet, over all timescales, the diffusion equation is not accurate and particle movement is more precisely given by the Langevin equations [6], which add a random force representing thermal fluctuations. It belongs to the family of Langevin dynamics which is a more complex version of Brownian dynamics, the latter corresponding to the limit where no average acceleration takes place (this approximation is also known as overdamped Langevin dynamics or as Langevin dynamics without inertia).

We define a "long" timescale as follows. As time increases the above distribution for the probability of finding a particle at position x becomes much broader, as the variance of the distribution grows with time, $\sigma^2 = 2Dt$. This means that the particle can be found at a much larger distance from its initial position. This occurs when the time of evolution of our system is much larger than the characteristic time of the diffusion, mainly: $t \gg \frac{L^2}{D}$, where L is the characteristic length of the system and D the diffusion coefficient.

Let us build our Brownian system. For that we will have n particles constituting our system, each with mass m and coordinates $r = r(t)$ [15]. The Langevin equation is given by:

$$\dot{r}(t) = v(t) \tag{11}$$

$$m\dot{v}(t) = -\lambda v(t) + \eta(t) \tag{12}$$

where r(t) is the position, v(t) the velocity of the particle, λ the damping coefficient and $\eta(t)$ a term introducing a stochastic noise which models the collisions of particles in

the fluid. As already announced it has a Gaussian distribution centered around 0 (thus, its mean value being 0) and a correlation function given by $< \eta(t1), \eta(t2) >= 2\lambda k_B T \delta(t-t')$, where k_B is the Boltzmann constant and T the temperature. The fact that the term $\delta(t-t')$ is introduced means that there exists no correlation between forces amongst different times, that is, there is no memory in the system. This is an important property of some stochastic systems (in the literature these are regarded as "Markov processes"), which also holds for Brownian motion. This is not completely true though; there exists some correlation between the microforces acting in different timespans upon the particles and leading them to move randomly. Nevertheless, these correlations are negligible and can be disregarded, as the description of the Langevin equation only focuses on long time-scales of "macroscopic" particles. In this scenario those forces cancel each other out, treating the forces as uncorrelated or "white noise".

From the previous assumption we can then regard the force upon particles as a stochastic memoryless variable, having no correlation between previous and following values. However a problem now arises: to integrate the Langevin equation we need to know the force acting upon the particle at each time, and as the force is a random variable without memory we cannot integrate it directly. This is a result of the irregularity of random variables, which are not differentiable and can not be expanded through a Taylor series. To incorporate this effect we feel the urge to define a new stochastic variable, ΔW (representing the integral of the white noise term over Δt), which shows the cumulative effect of the random force, that is: the impulse upon the particle. Thus, it is clear that this new variable can be written as the integral of the random force in our system [15], which we denoted by $\eta(t)$, obtaining:

$$\Delta W_i = \int_{t_i}^{t_i+\Delta t} \eta(t')dt' \tag{13}$$

and in substituting this new variable in the equations we previously derived we would obtain the following relations for the position and velocity of the particles:

$$r_{i+1} = r_i + \int_{t_i}^{t_{i+1}} dv(t) \approx r_i + v_i \Delta t \tag{14}$$

$$v_{i+1} = v_i - \frac{\lambda}{m}\int_{t_i}^{t_{i+1}} dv(t) + \frac{1}{m}\int_{t_i}^{t_{i+1}} d\eta(t) = \left(1 - \frac{\lambda}{m}\Delta t\right) v_i + \frac{1}{m}\Delta W_i \tag{15}$$

This result adheres correctly to the theoretical results of the Langevin equations found in the literature. However, we still need to introduce another factor to account for the drift velocity, which refers to the force acting upon the particles which is not due to random effects but rather to deterministic forces such as friction, fluid flow or other long-timescale effects (denoted by F(t)). This leads to the description of the medium as time-dependent, where the fluid flow affects the movements of particles. We then say that the fluid is non stationary and the Langevin equation must be rewritten to account for this new force. Therefore the equation in (12) becomes:

$$m\dot{v}(t) = -\lambda v(t) + \eta(t) + F(t) \tag{16}$$

This being our new equation, the position equation's discretized form remains $r_{i+1} = r_i + v_i \Delta t$, but the velocity of the particle must be rewritten in its discretized form as:

$$v_{i+1} = \left(1 - \frac{\lambda}{m}\Delta t\right) v_i + \frac{1}{m}\Delta W_i + \frac{F_i}{m}\Delta t \tag{17}$$

A simple dimensional analysis proves this equation is correct, as we get velocity components in both sides of the relation. In implementing this numerical system we would obtain the value of the velocity and position of the particle in each iteration:

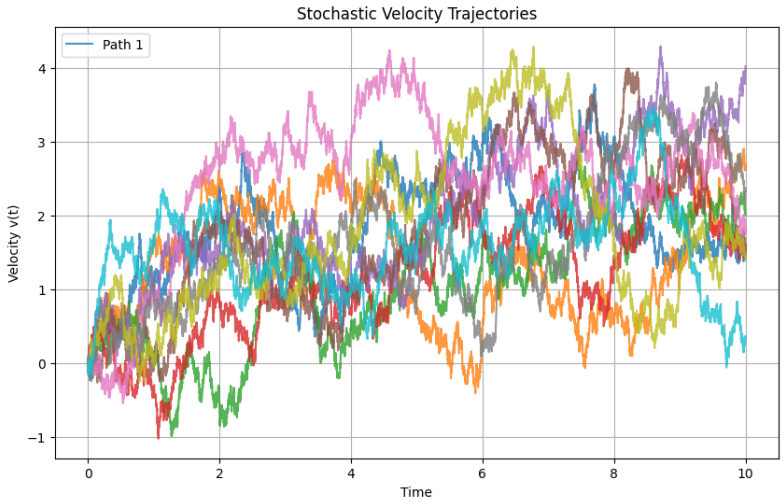

Figure 5: Simulated velocity trajectories (v(t)) of multiple particles undergoing Brownian motion with an external force, illustrating their stochastic nature over time

This deduction for the equation of motion of the particles is correct and in accordance with theory as time evolves to infinity. For the mathematical validity of these arguments see the extended work or the paper carried out by Naia [15].

So far, various terms have been included in our equation of motion and yet, one key component is still missing; interaction amongst particles. The first step in defining this force is to know the potential acting between the constituents of the fluid, which we will denote by $U(x_1, x_2, ...)$ which will describe how forces between particles influence their random motion.

To deduce this potential we can consider particles as nonpolar molecules that contain equal positive and negative electromagnetic charges with no net charge or dipole moment. As the molecules get closer (but still at "large" distances), the charges polarize and their null dipole moments start becoming nonzero. Therefore the interaction potential can be defined to high accuracy by $V(r) = -2E * d$, where $E = E(r)$ is the average electric field felt by one of the molecules when the second one is located at r, and $d = d(r)$ is the induced dipole moment. Let us consider a linear relationship between E and d. Solving Maxwell's equations (or using the multipole expansion of the Coulomb potential for a localized charge distribution), one finds that the far-field electric field of this dipole is proportional to :

$$\mathbf{E}(\mathbf{r}) = \frac{3\left(\mathbf{p}\cdot\hat{\mathbf{r}}\right)\hat{\mathbf{r}} - \mathbf{p}}{4\pi\epsilon_0 r^3} \tag{18}$$

So it is clear that the interaction potential V(r) is proportional to $-\frac{1}{r^6}$, which gives the commonly known Van der Waals potential. However, when distances are intermediate or small, polarization becomes stronger and induced dipole moments acquire more importance. To model these effects, including short-range repulsion, we use a phenomenological term that is equal to the square of the Van der Waals term, obtaining what we know as the Lennard-Jones potential. If we sum over all possible particles and derive by r we obtain the applied force on the system [4]:

$$\vec{F}_i = \sum_{r_{ij}<2\sigma, j\neq i} \left[-48\epsilon \left(\frac{\sigma}{r_{ij}} \right)^{12} + 24\epsilon \left(\frac{\sigma}{r_{ij}} \right)^{6} \right] \frac{\vec{r}_{ij}}{r_{ij}^2} \tag{19}$$

where we have taken $\vec{r}_{ij} = \vec{r}_j - \vec{r}_i$ and σ the sum of the potential to be over those interaction at a distance smaller than the cutoff, e.g., 2σ.

Including this result in the equations for the particle movements would give a very accurate description of the real Brownian movement.

Thus, the most important insight of this section is to observe how we have made use of the stochastic nature of particles in a fluid to describe their motion (along with some other considerations).

3.3 Chapter conclusions

It is clear by now, then, that stochastic modeling is of paramount importance in the description of various physical problems, such as classical/quantum noise and Brownian motion. Yet, stochastic processes are of the utmost importance not only in physical systems but also in economic systems. Now, then, we will focus on the task of describing financial systems by means of stochastic models as the ones described, and try to deduce the weaknesses of these methods. We will introduce the Efficient Market Hypothesis (EMH) and its weaknesses, and we will present a model that counteracts the EMH, showing that it is possible to achieve returns that exceed the market average through stock picking or market timing, in contrast to the EMH. Levy flights will be presented as a model that can be used to describe the behavior of financial markets and in contrast with the commonly used Brownian motion. From there the general properties of markets and financial time series will be developed.

4 Stochastic models in finance

Stochasticity has long been used to model asset prices and market trends. The most basic implementation of this is through the use of the geometric Brownian motion, where stock prices are modeled following the basic rules of Brownian motion we already introduced. The general formula for such models applied to asset prices yields [6]:

$$dS_t = \mu S_t\, dt + \sigma S_t\, dW_t \tag{20}$$

where S_t is the stock price at time t, μ the expected return (drift) , σ the volatility of the stock and dW_t is an increment of a Wiener process, accounting for the random fluctuations apparent in Brownian motion. We can define the drift term as $a(S_t, t) = \mu S_t$ and the diffusion term coefficient as $b(S_t, t) = \sigma S_t$ to simplify matters. To solve this equation numerically, we can discretize it over N timesteps of length $\Delta t = T/N$.The next step in the stochastic process can be given by the integral form: [6]:

$$S_{t_{n+1}} = S_{t_n} + \int_{t_n}^{t_{n+1}} a\big(S(x), x\big)\, \mathrm{d}x + \int_{t_n}^{t_{n+1}} b\big(S(x), x\big)\, \mathrm{d}W_x \tag{21}$$

Various methods are available to integrate this equation and obtain the expression for the next step. For a deterministic ODE, one of the most common methods is the Euler's method, which updates the solution by adding the product of the derivative (or drift term) and the time-step [1]:

$$x_{n+1} = x_n + f(x_n, t_n)\, \Delta t \tag{22}$$

However, our system is subject to random disturbances and so to account for this we need to make use of Maruyama's approximations;

$$\int_{t_n}^{t_{n+1}} a(S_x, x)\, dx \;\approx\; a\big(S_{t_n}, t_n\big)\, \Delta t, \tag{23}$$

$$\int_{t_n}^{t_{n+1}} b(S_x, x)\, dW_x \;\approx\; b\big(S_{t_n}, t_n\big)\, \Delta W_n, \tag{24}$$

where $\Delta W_n = W_{t_{n+1}} - W_{t_n} \sim \mathcal{N}(0, \Delta t)$ and $\Delta t = t_{n+1} - t_n$. Using these approximations, then, the updated step becomes:

$$S_{t_{n+1}} \;=\; S_{t_n} \;+\; a\big(S_{t_n}, t_n\big)\, \Delta t \;+\; b\big(S_{t_n}, t_n\big)\, \Delta W_n. \tag{25}$$

This simplification is known as the Euler–Maruyama method. Maruyama's insight was to modify Euler's method to also incorporate the effect of the stochastic term $b(S_x, x)dW_x$. Instead of just taking a step based on the drift, the update now also includes a term to account for the random shock.

An alternative approach would be to use the Milstein's method. Although it will not be mathematically developed in full detail here, we know that the step can be numerically implemented by [6]:

$$S_{t_{n+1}} \;=\; S_{t_n} \;+\; a\big(S_{t_n}\big)\, \Delta t \;+\; b\big(S_{t_n}\big)\, \Delta W_n + \frac{1}{2} b'(S_{t_n}) b(S_{t_n})(\Delta W_n^2 - \Delta t) \tag{26}$$

Note that the Euler-Maruyama result could have been easily obtained by setting the derivative of the variable b to 0 in the Milstein approach. Thus, the latter can be seen

as an enhancement of the simpler Euler–Maruyama method by incorporating additional terms from Itô's formula, which allow it to achieve a higher order of strong convergence when simulating SDEs. It does so by accounting for the effect of the derivative of the diffusion coefficient. This can be implemented numerically to see how a theoretical stochastic system behaves. Note that the difference between using EM or Milestein is negligible even for long time periods, as shown in the second graph:

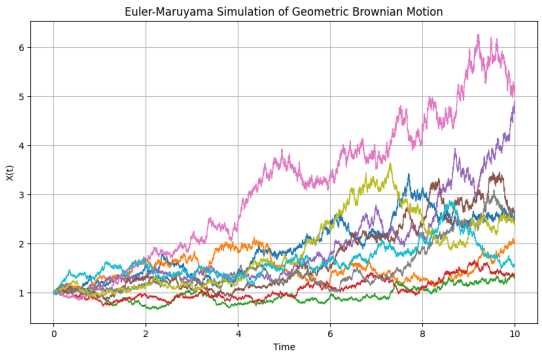

Figure 6: Numerical simulation of Geometric Brownian Motion using the Euler-Maruyama method, showing multiple stochastic paths for an asset price $X(t)$ over time

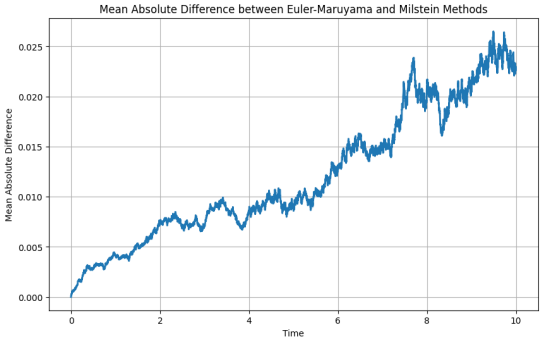

Figure 7: Mean Absolute Difference between simulations using the Euler-Maruyama and Milestein methods for GBM, illustrating the divergence over time.

Also note that these results are very similar to the ones we obtained in the Brownian motion section. The differences follow under the premise that the constants are not adapted equally for both scenarios.

Pure stochastic models, however, lack the ability to predict various scenarios and different models need to be used in those cases. In this work we will try to address the case where the market shows large jumps in the prices which cannot be modeled with regular Brownian equations. We will also give various reasons for rejecting the use of random processes to model finance assets by analyzing the general properties of financial time series. All of this will be done under the premise that, as it happens in physical systems, stochastic processes are key in financial systems but they also need to be processed and twisted accordingly to fit each scenario and describe its properties precisely.

4.1 The EMH and its weaknesses

Assuming the market is modeled with a motion equivalent to Brownian motion [3] is equivalent to saying that the market follows a hypothesis called the "Efficient Market Hypothesis" [8]. The latter is a financial theory that states that asset prices fully reflect all available information at any given time. This leads to a straightforward conclusion that it is impossible to beat the market (on the long run) since market prices should only react to new information. We refer to this as a market where no arbitrage exists. As the world becomes more globalized the idea that all the information is known by all agents in a market is not misconceived but could be seen as overly simplistic or idealistic; companies hold secrets, patent protection policies exist which prevent other parties from using certain information etc.

It might be necessary to make a key sidenote: while rejecting the EMH implies rejecting the description of financial markets with a Brownian motion, the inverse case does not necessarily hold, as a market can be regarded as efficient while returns show a different distribution rather than Brownian-like. Our discussion then focuses on trying to deny the EMH so that we can argue that Brownian motion is not the correct model to depict financial stocks.

The assumption that markets are random and follow the EMH (and that they can be described through Brownian motion), although sometimes accepted, is nowhere near a fundamental law. A few examples to counter this premise are being presented, guided by many decades of empirical research on return predictability. Thomas Lux and Michele Marchesi proposed that prices observed in financial markets frequently exhibit universal characteristics and patterns [12] . These behaviors are remarkably similar to scaling laws commonly found in physical systems characterized by complex dynamics. Such systems typically involve a large number of individual components whose interactions collectively produce emergent phenomena. This would prove wrong the EMH, which says that scaling in the return of prices would simply reflect the same scaling shown in the input of price information. In their model two types of agents were presented: the fundamentalists, followers of the EMH, and noise traders, who believe that the price does not immediately reflect all the information available in the market so they speculate what the real value should be based on their personal feelings or information.

The paper shows that investors with varied beliefs and strategies lead to markets following power-law distributions to which we commonly refer to as examples of self-organized criticality: this suggests that financial markets evolve towards a critical state where a small change could have huge impact on the market. The Brownian motion does not account for these apparent power-law distributions where extreme events are more common, so other models should be discussed to better understand market behavior if the EMH is not to be accepted.

To show that a market can indeed reject the principles of the EMH we will simulate a very simple yet eye-opening market. For that we will have both fundamentalist and noise traders actively taking part in our model. That way the fundamentalists will buy when prices are below a fundamental value and sell if they are above, while noise traders will trade randomly based on their beliefs. The price will be updated following this relation:

$$P_t = P_{t-1} + \alpha(F - P_{t-1}) + \sigma \cdot \text{random noise} \tag{27}$$

where P_t is the price at time t, F is the fundamental value, α is the speed of adjustment of the fundamentalist and σ is the volatility from noise traders. That way the price will be affected by both the demand from fundamentalists and from noise traders.

We begin our analysis from a market price of 100 arbitrary units and we set the fundamental price to 115; i.e, the price fundamentalists think of as "fair" [Figure 8]. With that in mind we let our system evolve for an time period of 200 steps, where each step we update the market prices. Clearly prices tend to the fundamental value rapidly but still show large fluctuations.

If we introduce yet another agent [Figure 9], called a trend follower, the behavior is

still different from the predictions of the EMH. These agents follow the trend trading strategy according to which they buy an asset when its price trend goes up, and sell when its trend goes down, expecting price movements to continue.

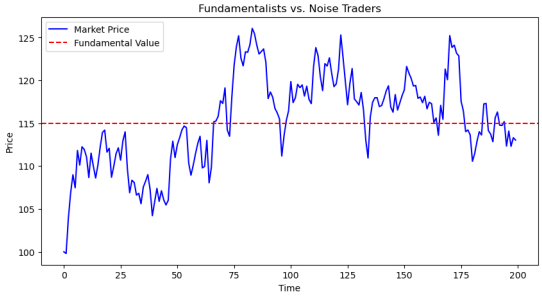

Figure 8: Artificial market simulation with two types of agents: fundamentalist and noise traders. Market Price (blue) fluctuates around the Fundamental Value (red dashed line). Parameters were adjusted as follows: Number of timesteps, initial price, fundamental value, speed of adjustment for fundamentalist, noise trader volatility = 200,100,115,0.1,2

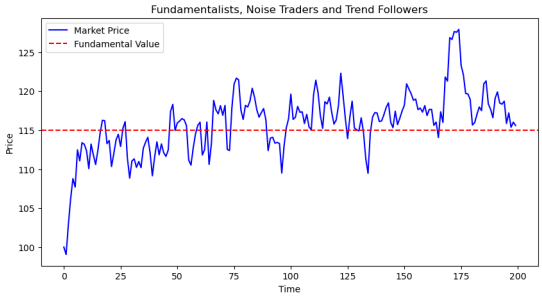

Figure 9: Artificial market simulation with three types of agents: fundamentalist, noise traders and trend followers.Market Price (blue) dynamics around the Fundamental Value (red dashed line). Parameters were adjusted as follows: Number of timesteps, initial price, fundamental value, speed of adjustment for fundamentalist, noise trader volatility,trend followers strength, adaptive agents' weight update factor = 200,100,115,0.1,2,0.05,0.02

As clearly illustrated in both figures, prices rapidly attempt to settle around the fundamental value. However, once this stabilization occurs, prices still deviate from the fundamental value, allowing the possibility of generating returns exceeding market averages through stock selection or timing strategies, which directly contradicts the Efficient Market Hypothesis (EMH). Although this market model is overly simplistic and lacks real-world complexity, it effectively captures the intrinsic market dynamics that challenge the validity of the EMH. Consequently, if the EMH does not hold completely, it becomes inappropriate to model markets using standard Brownian motion. Market inefficiencies such as investor biases, information asymmetries, or behavioral influences prevent asset price movements from strictly adhering to Brownian motion characteristics.

Therefore, in this section we have argued that the EMH may need to be rejected in some scenarios and thus another type of motion rather than Brownian motion needs to be found to describe the properties of these type of markets. However, before looking for an alternative, it might be illustrative to extract information about the nature of real financial time series so that we get an idea of what type of model better fits those properties.

4.2 Introduction to real financial time series

Empirical findings demonstrate that log-return distributions in finance frequently exhibit heavy tails that follow a power-law pattern. In contrast, a standard random walk — and by extension, Brownian motion — produces normal (Gaussian) distributions. To illustrate this, we can analyze the binomial model (a straightforward example of a random walk) as an analogue for a Brownian-like market. Its probability distribution is given by:

$$p(n) \approx \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(n-\mu)^2}{2\sigma^2}} \tag{28}$$

These are the type of distributions we get from Brownian systems. Obtaining the log returns we get the following graph:

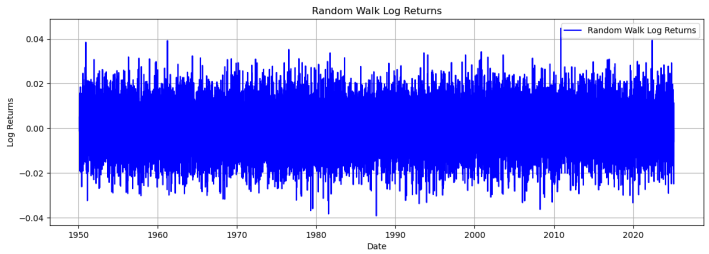

Figure 10: Simulated log returns from a random walk model, exhibiting fluctuations consistent with a Gaussian distribution.

Yet, as we have already mentioned, empirical data usually shows larger fluctuations. To see if this is true we will take some of the most influential stocks (mainly the SP 500, Euro Stoxx 50, Nikkei 225 and HSI) and plot their historical log returns for sufficiently large time periods.

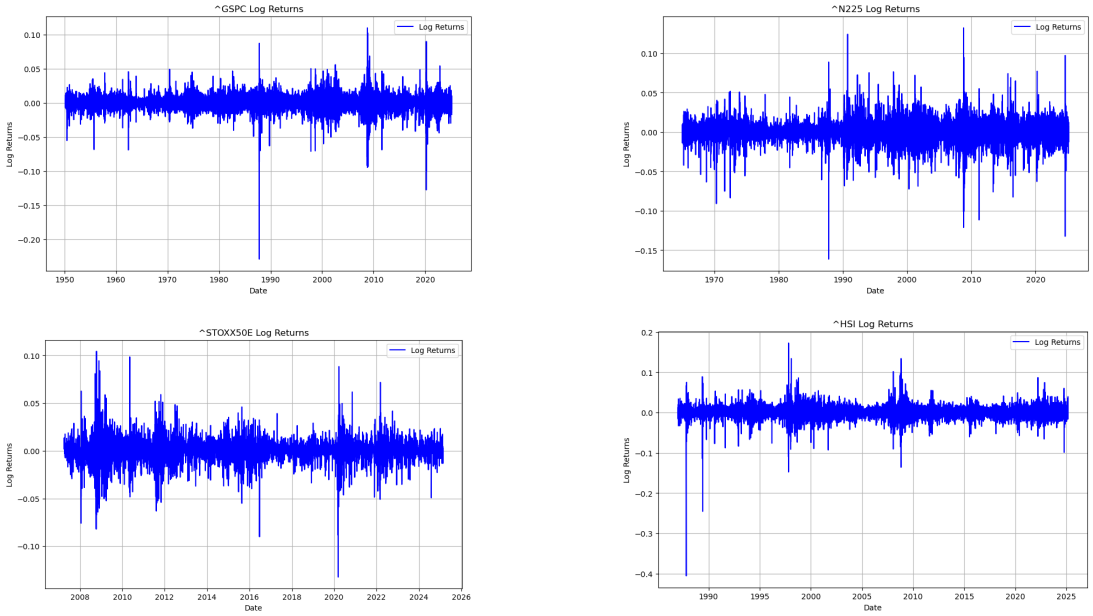

Figure 11: Historical log returns for various indexes: SP500, N225, EURO and HSI.

The results are obviously at odds with the previous random walk graph. By comparing these results, we see that the real stocks show fat tails (meaning larger deviations are more probable), while random walks do not resemble this behavior. If we were to obtain the log-return distribution of all the stocks, we would see a more peaked function than a Gaussian (remember, a random walk with steps following a normal distribution tends to a Gaussian distribution). In the literature, this is denoted by leptokurtosis.

It is clear by now that we need to employ an alternative stochastic process better suited to modeling phenomena that deviate significantly from normality. One such candidate is the Lévy flight, characterized by heavy-tailed distributions and long-range correlations. Lévy flights exhibit step sizes with occasional large deviations interspersed with many smaller steps, making them particularly useful for modeling extreme and unpredictable events.

4.3 Levy flights

Mathematically we say that Levy flights are another type of random walks that follow step-lengths with stable distributions. A distribution, $p(x)$ is said to be stable if, when taking samples from that distribution that have no correlation with each other, the sum of these also follow the same distribution. Let X and Y be two variables computed from the distribution $p(x)$. Then, we say $p(x)$ is stable if

$$Z = \eta X + \mu Y \tag{29}$$

has also the same distribution, for whatever values η and μ have. Thus, the linear combination of any amount of variables that follow this property would result in the same kind of distribution. It is important to note that there might be some changes in the form of the distribution, so identicality is not guaranteed. That means that it is prone to translation changes, where the new distribution can be centered in a different position than the original one, and to scale changes, where the new distribution's width can be stretched or compressed.

Intuitively the Gaussian distribution is stable, and so are all the distributions that fall under this entire class of (alpha-)stable distributions, whose characteristic function $\varphi(t)$ is given by:

$$\ln \varphi(t) = \begin{cases} i\theta t - \lambda|t|^{\alpha}\left[1 - i\beta\frac{t}{|t|}\tan\frac{\pi\alpha}{2}\right], & \alpha \neq 1 \\ i\theta t - \lambda|t|\left[1 + i\beta\frac{t}{|t|}\frac{2}{\pi}\ln|t|\right], & \alpha = 1 \end{cases} \tag{30}$$

Although the expression is quite complex, the parameters that concern us are α and β. When $\alpha = 2$ (and $\beta = 0$), we obtain the Gaussian distribution. If $\alpha < 2$, we have power-law tails with potentially infinite variance.

It may be worth noting that usually the term Levy flight is used more loosely to address those distributions that show a higher tendency to obtain values larger than average, that is, to refer to heavy-tailed random walks. In this sense they are not required to be alpha-stable distributions. It is in this regard that we will be using the "not so strict" definition of Levy flights from now on, meaning that our models can have steps that do not necessarily have stable distributions.

With that issue being tackled, we will apply the model to financial markets. As we have previously done with random walks, we will now simulate the price returns we would obtain by following Levy flights. For our model we will take into account a step-function which follows a power law Pareto distribution, very often used to describe inequalities such as the distribution of wealth in society: it has heavy tails, meaning extreme events

are more probable than in a Gaussian distribution. Its probability density function is given by:

$$f(x; x_m, \alpha) = \begin{cases} \frac{\alpha x_m^\alpha}{x^{\alpha+1}}, & x \geq x_m, \\ 0, & x < x_m. \end{cases} \tag{31}$$

where x is the variable(e.g., step size) we are evaluating, x_m is the minimum value x can take for the distribution to be well defined (thus, setting the lower bound) and α is the parameter that controls the shape of the tails of the distribution. When we introduced stable distributions we noticed that when α was 2 we obtained Gaussian distributions, so the variance was finite and tails were less heavy. Conversely, if α is smaller than 2, then the variance will be infinite (which also applies for Pareto distributions with $\alpha < 2$ for variance), and we will see extreme events more often. This can be clearly seen in the graphs shown below:

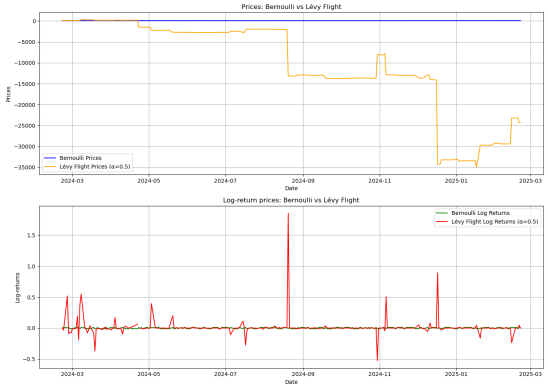

Figure 12: Simulated asset prices and log-returns for Bernoulli (Random Walk) steps (blue in top graph) and Lévy Flight steps with Pareto distribution($\alpha = 0.5$) (yellow in top graph) . Top: Price paths. Bottom: Log-return series.

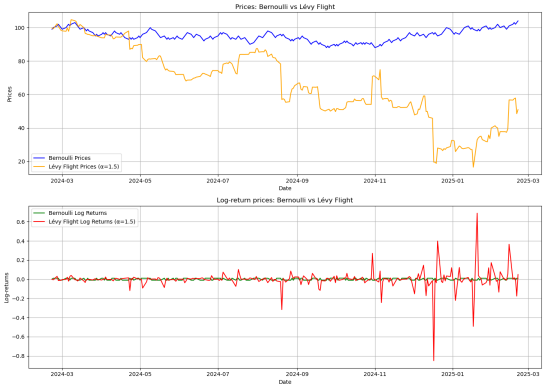

Figure 13: Simulated asset prices and log-returns for Bernoulli (Random Walk) steps (blue in top graph) and Lévy Flight steps with Pareto distribution($\alpha = 1.5$) (yellow in top graph) . Top: Price paths. Bottom: Log-return series.

As α increases from 0.5 to 2 larger jumps in the steps tend to occur less often, so the Levy flight returns do not show that many extreme values. To see how the returns of prices computed with Levy flights show this behavior, we will graph the difference between Levy stable distribution returns and Gaussian distributed returns as alpha changes values form 1 to 2. (Figure 14)

We have omitted values of alpha smaller than 1 because the jumps could lead to price differences so big that the plot would not visualize the effects needed.

It is clear that the smaller the value of alpha the greater the difference in value, as bigger jumps are more prone to happen, making the return differences more abrupt (recall that Gaussian systems do not have jumps at all) .We also see that even when alpha is precisely 2 the difference does not yield 0, as we should expect from Levy stable distributions(note that when $\alpha = 2$ Levy distributions become Gaussian). This is because the distribution considered is indeed, not alpha-stable, but a Pareto distribution with heavy

tails as we already discussed.

These behave like power tails, $x^{-(\alpha+1)}$, while Gaussians decay exponentially, e^{-x^2}. So even for high numbers of α we still see a difference in the subtraction of returns: No matter how large α gets, a Pareto's tail is never exponential, it remains a power law — just a steeper one [12]. That's still heavier-tailed than a normal distribution. So, while increasing α makes the Pareto distribution look "less extreme," it never really transitions into a Gaussian distribution. Instead, it just gets a thinner tail while staying skewed.

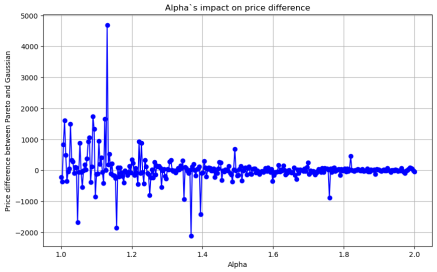

Figure 14: Impact of Pareto parameter α on the difference between simulated prices derived from Pareto-distributed steps (Lévy-like) and Gaussian-distributed steps (RW-like). Shows the price difference as a function of α.

We also see that, for values of α less than 2 the log-returns obtained with the Pareto distribution are similar to the ones obtained in the SP500 (and the other stocks presented above):

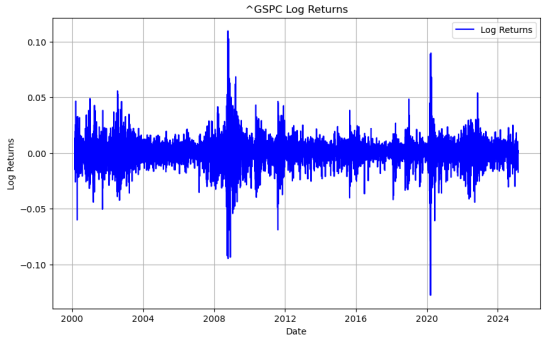

Figure 15: Empirical SP500 log returns between the years 2000 and 2024

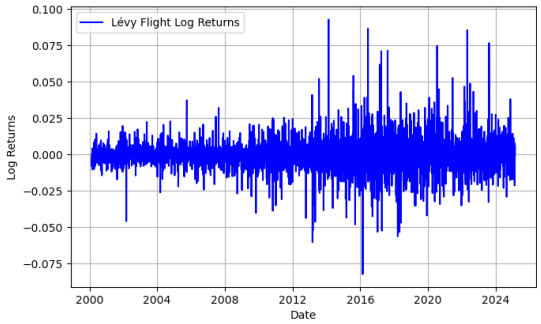

Figure 16: Simulated Lévy-like log returns for the same time period as Figure 15.

We then can say that markets are better described for values of alpha smaller than 2. This is the most important conclusion, as we have found a mathematical tool for describing the overall behavior of markets, mainly Levy flight type distributions.

Although the analysis carried out seems to fit well the data, this is only true for short-timescales or high frequency data. When we expand our analysis to higher time scales the results tend to revert to a Gaussian distribution. This is a known stylized fact in econophysics: the distributions of log returns for many financial assets have a tendency to form fat tails for high-frequency data. However, this tendency occurs solely for high-frequency data, but if the returns are obtained for longer time steps, this tendency disappears and the distribution approaches a Gaussian distribution. This behavior is known as aggregational Gaussianity [14]. To see if these arguments hold, we will carry out the same analysis we did before but with less frequent data

For this we have selected a Q-Q plot, a graphical method for comparing two probability distributions by plotting their quantiles against each other. The red line indicates a Gaussian distribution, while the blue dots are the data retrieved from the market. In these types of graphs a point (x, y) on the plot corresponds to one of the quantiles of the second distribution (y-coordinate) plotted against the same quantile of the first distribution (x-coordinate).

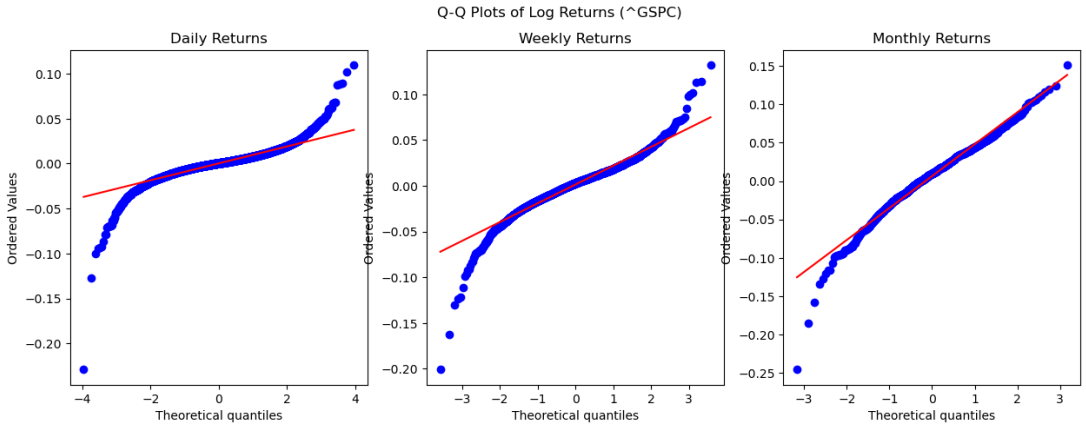

Figure 17: Q-Q plots of log returns for the SP 500 market (data from 1950 to 2025) against a standard Gaussian distribution. Plots are shown for daily, weekly and monthly return frequencies, illustrating aggregational Gaussianity.

Clearly, daily data shows more deviation from the Gaussian distribution, meaning jumps occur more often. On the contrary, as we increase the time scale, the data seems to reshape and collapse into a Gaussian distribution, as it can be clearly seen in the last graph.

So, what is the key takeaway point from here? Levy flights are often used in the context of modeling anomalous diffusion and certain types of random walks, particularly when the data exhibits heavy tails. In financial markets, for example, Levy flights have been used to model high-frequency data or short-term price movements, where there are more frequent, larger jumps than what would be expected under a normal (Gaussian) distribution. At shorter time scales, market returns or other types of data may exhibit these large jumps (or "fat tails") characteristic of Levy flight behavior. In this context, Levy flights can be a better fit for data because they allow for larger, more frequent deviations from the mean, which is often observed in high-frequency market data.

However, for longer time scales or low-frequency data (e.g., weekly or monthly returns), the distributions of returns often become more Gaussian in nature due to the central limit theorem, where the sum of many small, independent random variables tends toward a normal distribution. In this case, Brownian motion (which models Gaussian behavior) becomes a more appropriate model, as the larger jumps and extreme events are less frequent.

Therefore, even if Lévy flights generally provide a better framework to describe financial markets than Brownian motion, especially in capturing real-world features such as fat tails, extreme events, and discontinuities, they come with their own limitations. Thus, in

order to capture complex behaviors, a more developed model needs to be considered.

4.4 Do markets have memory?

It is important to underline that the EMH and, therefore, markets modeled by Brownian motion, show no dependency on past returns, meaning no memory correlation seems to appear. However, the failure of the EMH can lead to memory effects and long range dependence. Then, is there really memory in markets? [11] This is a nuanced aspect of financial markets and to give an answer we need to break it down a little bit further. To show if there is memory, we will examine the autocorrelation of log-returns:

$$G_r(T, \Delta t) \equiv \langle r(t, \Delta t) \cdot r(t + T, \Delta t) \rangle_t \tag{32}$$

Here we have assumed t to be the time when we evaluate the log-returns, Δt the time difference over which returns are calculated (e.g., daily,weekly),and T is the time lag or phase, so we evaluate the autocorrelation of returns calculated over Δt at time t with those at time $t+T$, averaging over all t . The result being 1 indicates perfect correlation, 0 no correlation at all and -1 perfect anti-correlation. Note that the term "auto" in "autocorrelation" only means that we compare the same time series with its lagged version, while correlation can be computed with different time series. For our analysis we will, for now, only be concerned with single time series.

Therefore, using the SP 500 as the target market, we have graphed the autocorrelation of its returns over a sufficiently large period using data from Yahoo finance. The results obtained are the following:

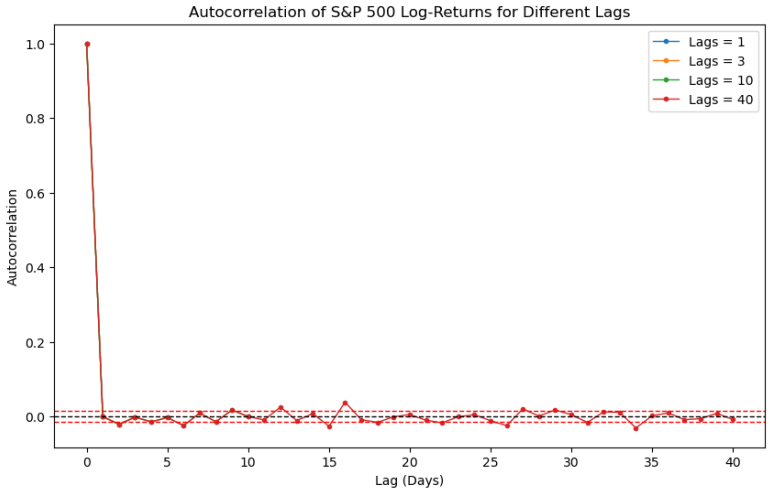

Figure 18: Autocorrelation of SP500 log-returns over the period 1980-2025 for different lags(1,3,10 and 40 days). The red dashed lines, drawn at $\frac{1.96}{\sqrt{(N)}}$, show the confidence interval to account for noise. We have taken lags of 1,3,10 and 40 trading days.

It is evident that there is little or no linear memory shown in the raw log-returns in the market, as the autocorrelation of log-returns decays very quickly to zero, often within a few lags. This distribution appears to have an exponential-like decay for very short lags

(e.g., form $e^{-\frac{T}{\eta}}$ where values of η up to 15 can be observed, meaning they fall of very rapidly) and aligns with the classical results of the EMH and the lack of memory dependence, thus confirming the idea that price movements resemble random walks. From the graph, we can also obtain that the greater the number of lags, the closer the autocorrelation is to zero (beyond statistical noise), showing yet less correlation.

A clever trick carried out in financial econometrics to show volatility clustering is to compute the squared or absolute returns. This falls under the premise that raw returns show both the magnitude and direction of prices, while in obtaining the squared values you focus solely on the size of the fluctuations, highlighting volatility. So, price changes could be completely random, but their magnitude might still exhibit persistence in time (i.e., large changes tend to be followed by large changes, and small by small), meaning trends in volatility would cluster in time. Figure 19 shows the absolute returns instead of raw returns, in the same market, for the same time period and with the same number of lags:

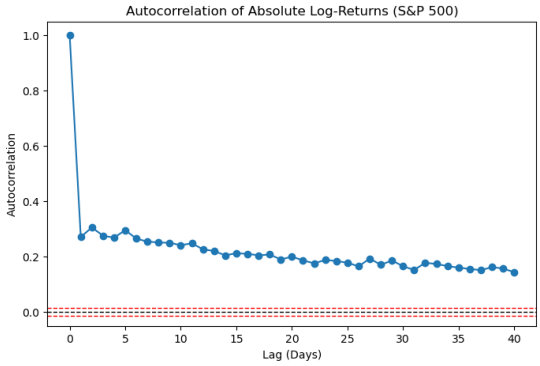

Figure 19: Autocorrelation of SP 500 absolute log-returns over the period 1980-2025. The red dashed lines, drawn in $\frac{1.96}{\sqrt{(N)}}$, show the confidence interval to account for noise. We have taken a maximum lag to of 40 trading days

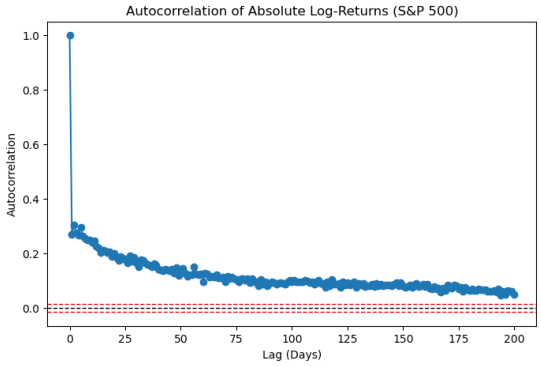

Figure 20: Autocorrelation of SP 500 absolute log-returns over the period 1980-2025, for lags up to 200 trading days. The red dashed lines, drawn in $\frac{1.96}{\sqrt{(N)}}$, show the confidence interval to account for noise.

Clearly there is a key difference from the raw log-returns: we now see a more power-law based tail (or at least a much slower decay), of the form $T^{-\eta}$, rather than an exponential decay we obtained for raw returns. This shows that there exists a correlation in the magnitude of returns: this slower decay indicates that volatility is not random but clusters over time.

We have already mentioned that increasing the number of lags would result in a higher term in the exponential function of the decay of raw returns, meaning even less correlation would be seen. In the absolute returns, though, this rapid decrease is not as noticeable, as the function now decreases like $T^{-\eta}$. Therefore, calculating the absolute log returns for a lag of 200 we obtain the results shown in Figure 20

Evidently, we can still see that the drop in autocorrelation is still not as steep as in the case of raw log-returns. Then, the main question still arises: is there memory in the market? We could argue that raw returns follow the EMH, meaning there is no memory and prices follow random patterns, with no predictive capabilities. However, based on

the fact that absolute returns have memory, we argue that volatility exhibits clustering and memory shows up in volatility. Therefore if a period of high/low volatility shows up in the market it is likely to persist in time due to this memory, even if individual price changes are independent. The answer then reduces to the type of memory we are dealing with.

4.5 Multiple financial time series

We have already implicitly come to the conclusion that various properties seem to appear in real financial time series, even if they are not derived from a profound theory. Nonetheless, until now we have focused on analyzing one asset or economic variable at a time, such as the price of a stock, a fairly simple model. This ignores interactions with other assets or markets and limits the broader market behavior. For this reason we feel the urge to look into several financial time series simultaneously. That way we will be able to capture interdependencies and correlations between assets and to carry out deeper analysis such as modern portfolio theory or other diversification models. Indeed, in real markets several stocks are traded at the same time and looking at single ones would mean neglecting their interactions, which is a bold move.

When we deal with relatively simple systems we can define their behavior by describing the interactions amongst all their constituents. However, when complexity starts to arise this becomes impossible. As an illustration we shall consider the description of a very complex physical system: the nucleus. Later on we will create an analogy with the complexity of financial markets.

4.5.1 Phenomenological models in physics and finance

In the study of nuclear interactions the simplest approach would be to get the interaction potential between two particles and generalize the results for the whole system. We can make use of the deuteron, the simplest nuclear model, made up of one proton and one neutron [4].If we analyze the z component of the magnetic moment of the deuteron and as we experimentally know that J=1, we conclude that the values the orbital angular momentum l can take are restricted to ${}^3S_1, {}^3D_1, {}^1P_1, {}^3P_1$. In order to get a magnetic moment in accordance with experiments and following the parity conservation principles, we need to have a wave function of the form $\psi(x,t) = c_S\psi_S(x,t) + c_D\psi_D(x,t)$, so terms corresponding to l=0 and l=2 appear. Thus, the Hamiltonian of this system needs to include a tensor term which mixes these two states. With all this in mind we get a general nuclei-nuclei potential of the form:

$$V(1,2) = V_c(r,p,L) + V_\sigma(r,p,L)\,\sigma_1{\cdot}\sigma_2 + V_T(r,p)\,S_{12}(r) + V_{LS}(r)\,L{\cdot}S + ... \tag{33}$$

This is the most general potential obeying invariance amongst particle exchange, translations, Galilean transformations, rotations, parity and temporal invariance (the term of the isospin still needs to be taken into account, but the structure will remain similar)

See that the generalization of this from two-bodies to the general picture of the nuclear systems is impossible, not mainly because of the vast number of particles involved but also

because the potential is very complex and cannot be clearly identified; it is an effective interaction. It is in this regard that we aim to create idealized models that capture part of the physics involved and explain the limited data available.

Clearly, then, to describe complex physical systems a general description needs to be developed. This is the same scenario we find in financial markets, where it is often impossible to capture and describe each component taking part. For this reason we need to create an "effective interaction" which models correctly the system without the need to delve into the intricacies of each all individual interactions. Also, note that there may exist various models, each valid for the description of certain properties, so there is no universal theory that can cover all the properties at the same time.

4.5.2 Correlation in financial markets

As a first step to look into the relationships among markets is to attain the cross-correlation values. It is important to remember that correlations show us the microscopic interactions among buyers and sellers (or rather, the macroscopic outcome of their aggregated decisions, reflected in price co-movements). The cross-correlation is very closely related with the covariance of two random vectors i and j, which in our case would refer to two different financial time series, and gives the similarity between those two variables. We can define the formula for two time series i and j by [9]:

$$\begin{aligned} C_{ij}(\Delta t) \equiv \frac{\langle r_i(t,\Delta t) r_j(t,\Delta t)\rangle_t - \langle r_i(t,\Delta t)\rangle_t \langle r_j(t,\Delta t)\rangle_t}{\sqrt{(\langle r_i^2(t,\Delta t)\rangle_t - \langle r_i(t,\Delta t)\rangle_t^2)\left(\langle r_j^2(t,\Delta t)\rangle_t - \langle r_j(t,\Delta t)\rangle_t^2\right)}} \\ \equiv \frac{Cov[i,j]}{\sqrt{Var[i]Var[j]}} \end{aligned} \tag{34}$$

this equality is known in the literature as the Pearson Correlation Coefficient [13], which measures the linear dependence of two sets. A value of 1 means perfect correlation, 0 means no correlation at all and -1 perfect negative correlation.

If we assume that each individual stock follows a random walk then the correlation matrix, C_{ij}, would be the identity matrix (for independent random walks), the same value we would obtain as time tends to infinity (for ensemble averages if they are truly independent). This follows from the fact that independence implies no correlation and thus it obeys for $i \neq j$:

$$\langle r_i(t,\Delta t) r_j(t,\Delta t)\rangle = \langle r_i(t,\Delta t)\rangle \langle r_j(t,\Delta t)\rangle \tag{35}$$

which if we put into equation (34) yields 0 for all off-diagonal elements but the ones with the same index values (diagonal elements which are 1 by definition of autocorrelation of a variable with itself, normalized).

For finite time series a symmetric distribution of the non-diagonal elements centered around 0 should be expected from purely random, uncorrelated series, in accordance with some aspects of Random Matrix Theory [16]. To see if this holds in finance, we will take the four markets already introduced in this work and carry out the correlations amongst

them, plotting a histogram to get visual results. Using SP 500, NIKKEI, HSE and EUROSTOXX50 we obtain the following :

$$C_{ij} = \begin{bmatrix} 1 & 0.248162 & 0.243312 & 0.647292 \\ 0.248162 & 1 & 0.499384 & 0.377193 \\ 0.243312 & 0.499384 & 1 & 0.390171 \\ 0.647292 & 0.377193 & 0.390171 & 1 \end{bmatrix}$$

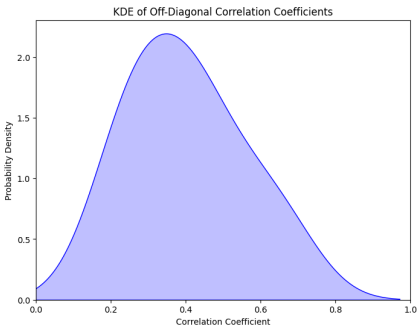

Figure 21: Kernel Density Estimate (KDE) of off-diagonal correlation coefficients from the 4 x 4 correlation matrix of SP 500, NIKKEI, HSE and EUROSTOXX50 indices.

Here, clearly, we see the average correlation being roughly around 0.4, clearly shifted towards positive values and in contrast to the null value we should expect from random time series. This means there exists correlation between these market indices, rejecting the idea they follow independent random walks.

However some might argue that the time series analyzed so far refer to indexes rather than individual stocks, which can introduce some issues. Firstly, the sample size is not large enough, as 4 indexes may not reflect the reality for the rest of the markets. Also, each index already aggregates the behavior of many stocks: the correlation between indexes tends to be higher as a result of this than that of individual stocks. For these reasons, we might want to switch our analysis to the biggest individual stocks rather than indexes, where we take data from 15 of the most relevant companies worldwide. Here too, in taking the KDF function we obtain the results of Figure 22.

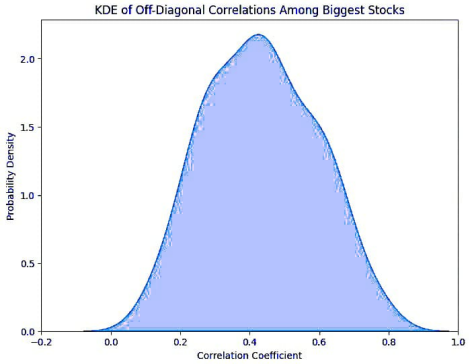

Figure 22: Kernel Density Estimate (KDE) of off-diagonal correlation coefficients among 15 major global stocks.

Similar results as those shown for indexes appear, where the shift in the correlation coefficient is arguably more positive, leading to more interconnection between these individual stocks than between the indices. But, is this possible? The answer could lie in the fact that most of the stocks taken into account here are part of the same demographic background - they are part of the US stock market - whereas previously we computed the correlation for indexes worldwide. A simple solution would be to carry out the same analysis for the most relevant stocks from all over the world and see if the positive correlation still persists.

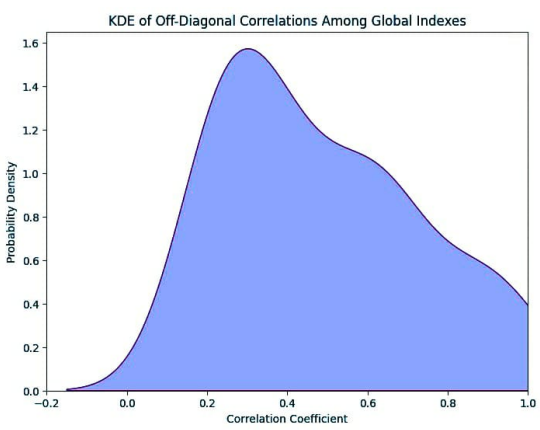

Figure 23: Kernel Density Estimation (KDE) of off-diagonal correlations among global stocks.

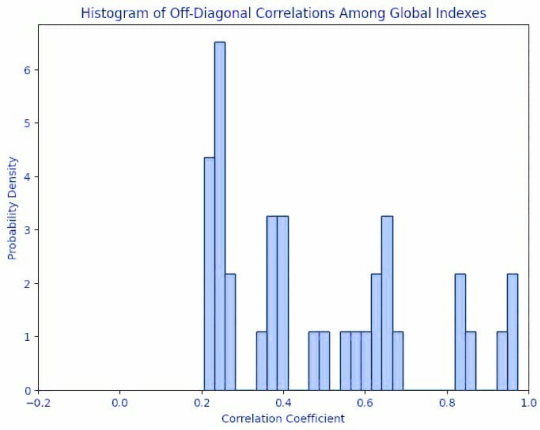

Figure 24: Histogram of off-diagonal correlations among global stocks.

Here we also see a positive change in the correlation values, but the peak has shifted somewhat towards lower positive values while at the same time becoming smaller in value (that is, a broader distribution) so we can argue that most stocks are less correlated than before, although still there is a clear positive correlation amongst them.

We also observe that the tail on the right side of the function is heavier than before leading to stocks that are very heavily correlated. Note that, along with the KDF, we also plotted the histogram to show that it is impossible to get correlation values greater than 1, even if the KDF distribution appears to show them (due to kernel smoothing effects at boundaries). This is only as a result of how the KDF functions are drawn and one of the main setbacks of it, but it can be directly corroborated by looking at the histogram which show that values than correlation values are always smaller than 1.

This analysis, then, has shown that markets are indeed interconnected, as positive correlations appear amongst them. Otherwise symmetric distributions around 0 should have been obtained.

We have also concluded that individual stocks show less average correlation than indexes do, although heavier tails in the correlation distribution also appear, leading to various highly correlated pairs of stocks. We have also seen that markets of the same demographic background/s tend to be more correlated. The major conclusion, then, is that it is necessary to consider the intricate interconnections amongst stocks and that neglecting them would only lead to misleading results, as their influence is clear.

4.5.3 Eigenvalues of the correlation matrix

We have already seen that correlation exists. So in account of the complexity of financial markets and in analogy with the description of complex physical systems, we need to compute our phenomenological theory for markets. Let us then proceed by extracting the eigenvalue spectrum of the correlation matrix ($C\phi_i = \lambda_i\phi_i$) [6]. These values help in finding noise in data: they show which are the different patterns that appear in the market, each pattern given by a different eigenvector, and how strong each pattern is, denoted by the eigenvalue of each eigenvector. Thus, in obtaining all eigenvectors and

eigenvalues in the market we can extract information such as common market movements (the "mode") or noise appearances, which belong to eigenvectors or patterns with low eigenvalues.

To show if the distribution of empirical eigenvalues follows a distribution different from that built with random time series, we will take our null hypothesis to be exactly that: we will compute the distribution of eigenvalues of a correlation matrix built from various random time series (which we will refer as our "artificial random markets") [6] and take this distribution as our baseline. Then we will take various stocks and produce the distribution for their eigenvalues, and we will compare if this function aligns with the null hypothesis. Our aim then will be that of showing that both distributions show differences to account for the non randomness of financial time series and to show that correlation exists, as we already proved by other methods.

We define n x n random matrices of the form $H = X \cdot X'$, where we have defined X to have entries which are random variables following a Gaussian distribution (mean 0, variance 1) and with dimensions of n x m (X' is its conjugate transpose) [6]. These will serve as the correlation matrices for our artificial model, for they represent data that is random and uncorrelated and accounting for artificial financial markets that are purely random. We have built a system with 100 uncorrelated time series (N=100 stocks) with 1000 time points (T=1000 days), yielding a distribution shown in Figure 25. This graph illustrates the theoretical high and low limits of the eigenvalues associated with a random variable covariance matrix. We have drawn both the histogram of empirical eigenvalues from one random matrix and the KDF (which should approximate the Marchenko-Pastur law [6]) for more visual results. See that most eigenvalues should lie in the range predicted by RMT (e.g., approximately [0.3,1.75] for Q=N/T=0.1 and $\sigma^2 = 1$).

We will also redo the same analysis with the correlation matrices from real stocks, and plot the results which are shown in Figure 26.

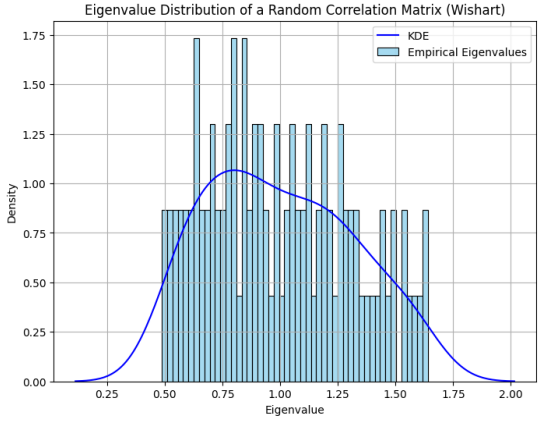

Figure 25: Theoretical eigenvalue spectrum for a random correlation matrix from a system composed of 100 artificial random time series of 1000 points each.

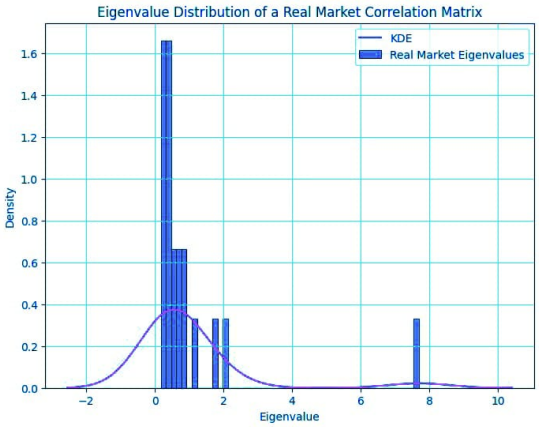

Figure 26: Eigenvalue spectrum (histogram and KDE) for real markets, using stocks from worldwide markets (not only USA).

Evidently both graphs differ, for the latter shows some eigenvalues (mainly, one) much bigger than the previous ones found in the random matrix case. However, most of the

eigenvalues fall under a distribution which looks similar in both cases (the "bulk"), although the bulk of eigenvalues from real data is narrower or shaped differently than the RMT prediction for purely random data. These bulk values largely account for random co-movements or statistical noise and they are often filtered out in RMT-based analyses as they may not reflect the real structure of markets. It is in this regard that we say that eigenvalue analysis helps in extracting noise from data!

The largest eigenvalue standing out typically refers to the common movement shared by all stocks, which is denoted as the market trend or market mode. An example of this could be market-wide shocks like COVID, the Russia-Ukraine war or else. We can also see other eigenvalues (e.g., around 2 in the figure) which differ from the bulk of the theoretical function and, therefore, these account for sectoral behaviors in stocks as well as geographical factors or other issues such as inflation sensitive stocks.

All these properties are neatly deduced by looking, as well as the eigenvalues, to their corresponding eigenvectors [16]. In the graphs below we have plotted the components of different representative eigenvectors:

- For the eigenvector corresponding to the eigenvalue of 7.7 we see that the components are more or less homogeneous (all positive and of comparable magnitude), indicating that most of the stock participate equally in that eigenvector; this, thus, leads to the property we already announced which says that large eigenvalues can lead to a market "mode". This shows that most markets tend to act together and are usually due to large-scale events affecting all stocks. See that all of them show positive component values, meaning they all move in the same direction under this type of events.

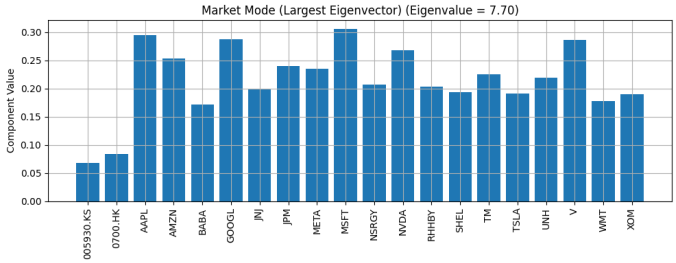

Figure 27: Components of the eigenvector corresponding to the largest eigenvalue (Market Mode, $\lambda \approx 7.7$. Stock tickers are on the x-axis.

- For a mid-range eigenvalue (e.g., $\lambda \approx 1.99$) we see stocks whose component contributions are considerable while other contribute almost nothing. This shows that sector behavior appears, where changes affecting a certain sector lead to only those stocks changing in value while the rest do not. In this case both negative and positive values appear, meaning contrary behaviors appear in different stocks. If we analyze closely this case we see that stocks like "AMZN","META","MSFT" or "NVDA", which are tech/consumer stocks, all have positive components while "JPM","WMT" or "XOM", which are financial/energy stocks, show negative components, meaning that this type of eigenvalue leads to increments in the technological stocks while

there are decays in financial and energy ones, showing correlation amongst stocks that share common features within these groups but opposition between them.

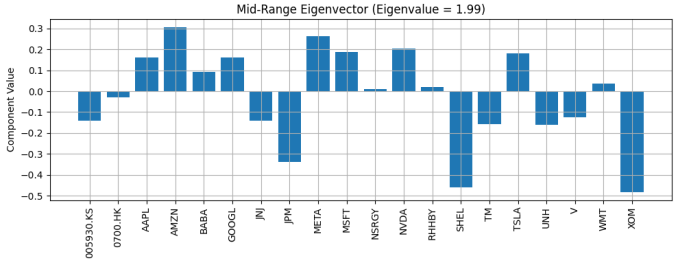

Figure 28: Components of an eigenvector corresponding to a mid-range eigenvalue (Sector Mode, ≈ 1.99).

- For low eigenvalues the components of the eigenvectors look fairly random and point to noise in the data, which are the ones we need to discard in our analysis. This is the main characteristic of this model; the ability to detect these patterns and discard them from the really valuable data.

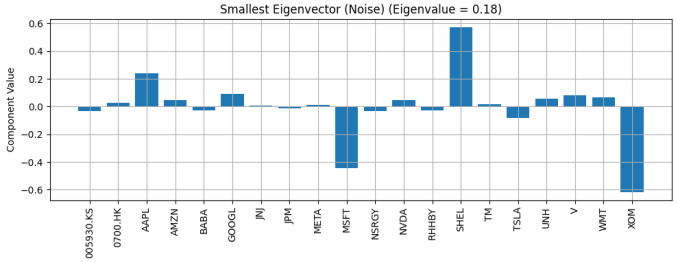

Figure 29: Components of an eigenvector corresponding to a small eigenvalue (Noise Mode, $\lambda \approx 0.18$).

If we analyze closely the "bulk" values or the ones corresponding to noise we can see that a special type of distribution seems to appear for the eigenvector components: the Porter-Thomas distribution [10]. Therefore a straightforward approach to know if the eigenvalues correspond to noisy data is to see if their eigenvector components follow that distribution, as it can be seen in the graph below:

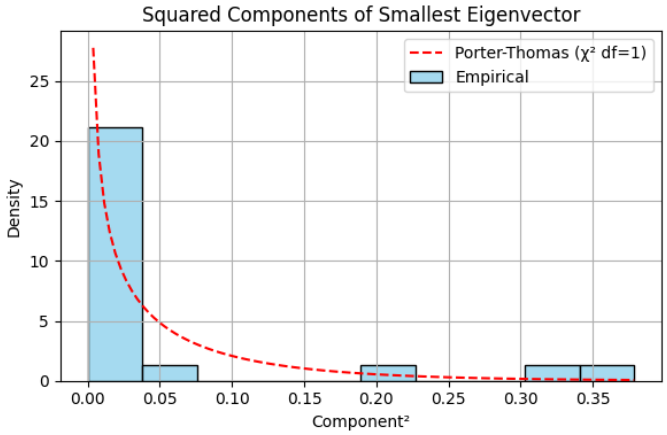

Figure 30: Distribution of the squared components of the smallest eigenvector (noise mode), compared with the theoretical Porter-Thomas distribution (χ^2 with 1 degree of freedom).

As it is clear real markets differ from the theoretical results of RMT for purely random matrices and show that there is indeed correlation across different stocks, in total accordance with the previous results from the last section. We observe that we find maximum eigenvalues significantly larger than those predicted by the upper bound of RMT's bulk. For small eigenvalues we obtain eigenvectors whose components are distributed according to theory. However, large eigenvalues give rise to eigenvectors which deviate significantly from such random-like component distribution. Note that each eigenvector component $\phi_i(j)$ tells how much stock j contributes to the corresponding eigenvalue mode. Thus, the eigenvalues of a correlation matrix separate meaningful market trends, sectoral structures, and random noise, making them a powerful tool in financial data analysis. This tool is commonly known in the literature as the Random Matrix Theory.

4.6 Correlation filtering

The processing task of a correlation matrix can be challenging due to the vast amount of information contained within this matrix, where all correlations across all pairings are included. Therefore, in building our phenomenological model, our goal now resides in finding a smaller subset of correlations which reflects the overall behavior of the original correlation matrix. For that we will make use of the conclusions drawn above and we will divide the correlation matrix into three subsets, one to give account for the overall market behavior, $C_{overall}$, another one for the individual sector behaviors, C_{sector} and the last one to address the noise and random fluctuations, C_{noise} [13]. Using an expansion in terms of eigenvalues λ_i and eigenvectors it reads

$$C = \lambda_0|\phi_0\rangle\langle\phi_0| + \sum_{i=1}^{k} \lambda_i|\phi_i\rangle\langle\phi_i| + \sum_{i=k+1}^{N-1} \lambda_i|\phi_i\rangle\langle\phi_i| \tag{36}$$

where the terms correspond to $C_{overall} + C_{sector} + C_{noise}$, respectively. We define k to be the number of eigenvalues between the bulk and the maximum value and N to be all the eigenvalues. Bear in mind that C_{sector} helps identify groups of stocks influenced by common factors.

Also some of the assumptions of the Random Matrix Theory are to be taken carefully as they might not always hold, i.e, the largest eigenvalue does not always show the market mode accurately. As an illustration lets assume we have a market composed of N stocks with correlation coefficients that are weak overall, but with a small group of k stocks that are very strongly correlated amongst themselves. We then could divide our correlation matrix conceptually:

$$C = C_{overall} + C_{sector} \tag{37}$$

In such a case, the largest eigenvalue might correspond to the strongly correlated sector rather than a market-wide mode if that sector's internal correlation is dominant enough.

By taking $S \subset \{1, \ldots, n\}$ as the set of indices of stocks belonging to a strongly correlated sector, we can define a simplified model for this part of the correlation matrix as [12]:

$$(C_{\text{sector}})_{ij} = \begin{cases} \rho & \text{if } i \neq j \text{ and } i, j \in S \\ 1 & \text{if } i = j \in S \\ 0 & \text{otherwise} \end{cases} \tag{38}$$

where we take ρ to be very large indicating very strong correlations amongst those sector markets. The results show as follows:

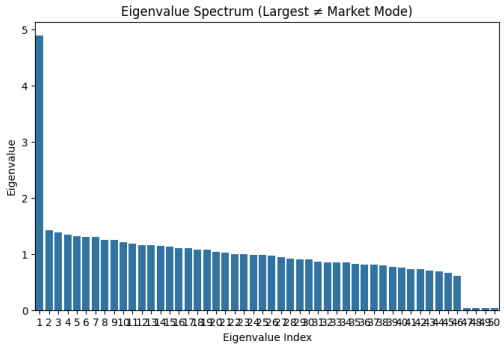

Figure 31: Eigenvalue spectrum of a model correlation matrix where a small sector of stocks has very high internal correlation, while other stocks are less correlated. The largest eigenvalue is dominated by this sector.

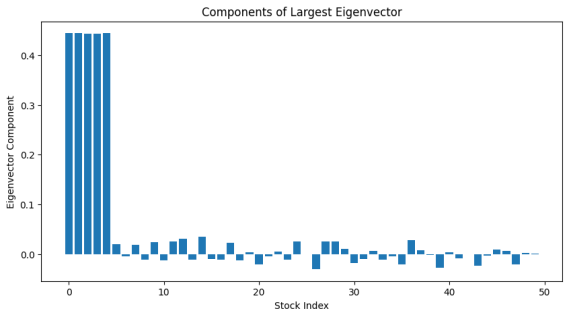

Figure 32: Components of the eigenvector corresponding to the largest eigenvalue from the model in Figure 31. Components are large only for stocks within the highly correlated sector, indicating a sectoral mode.

We know that usually the market mode obeys that:

$$\lambda_1 \gg \lambda_{\text{ohters}}, \quad |\phi_1\rangle \approx \text{market mode}$$

But when a sector dominates, as it is in the case we developed, it follows that:

$$\lambda_1 \to \text{sectoral mode}, \quad |\phi_1\rangle \approx \text{localized on sector}$$

Thus in this case the largest eigenvalue does not correspond to the market mode, as most markets do not contribute to it, but rather it reflects a sector where few components influence it. This illustrates that correlation filtering must be interpreted with care as the

largest eigenvalue is not always the market mode.

Also removing market or noise modes could lead to unmeaningful matrices and filtering could lead to improper correlation matrices requiring alternative constructions, so there is no unique way to remove market and noise modes, meaning different filtering approaches can yield different results. Thus, even if RMT is a useful filtering technique, it is not sufficient and additional processing is often required. Some techniques use RMT theory as an input for other processing tasks such as hierarchical clustering: a non-overlapping type of algorithm that groups objects into clusters without setting the number of clusters beforehand so that we can group stocks that have similar patterns.

4.6.1 Modularity optimization

One of the most famous methods to detect community structures within networks with non overlapping nodes is modularity optimization, developed by Mel MacMahon and Diego Garlaschelli [13]. It is a type of unsupervised clustering used to detect communities or clusters in a network by maximizing a measure called modularity, Q, which shows how well a network is divided into communities. For that it compares the actual number of edges within communities to a random null model where edges are placed at random and checks whether the connections amongst the real network nodes are more probable than the ones obtained with the null hypothesis. The modularity function reads as

$$Q = \frac{1}{2m}\sum_{i,j}(A_{ij} - < A_{ij} >)\delta(\mu_i, \mu_j) = \sum_{i=1}^{\mu}(e_{ii} - a_i^2) \tag{39}$$

where the first term is the normalization factor and the sum is over all pair of nodes, μ_i referring to the community to which node i belongs. The delta function ensures that only nodes of the same community sum to the function. A_{ij} (the adjacency matrix) yields 1 if there is a link between nodes i and j, and 0 otherwise. In this formulation we are aiming to maximize modularity. $< A_{ij} >$ is the null hypothesis: the expected number of edges between i and j if connections were random, which we aim to compare our model to, known as the configuration model. In this model, for undirected graphs, the expected number of edges between two nodes is usually given by this expression (or equitatively, the probability of nodes i and j having a link):

$$< A_{ij} >= \frac{k_i k_j}{2m} \tag{40}$$

where k_i and k_j are the degrees of the corresponding nodes (the number of edges incident to each node) and m is the total number of edges in the graph. Thus the equation for modularity is nothing more than the difference between the actual number of edges within communities and the one given by the null hypothesis, summed over all pairs of nodes within the same community. It is important to note that the accuracy of the results relies to a great extent on the null hypothesis selected, so the choice is of great importance. Our aim then will be towards seeking a proper null hypothesis to extract meaningful communities from markets, that is, to detect sectoral behaviors.

Then this type of analysis seems a good approach to detect groups in our model, where high correlations may appear amongst groups.

We could treat C_{ij} as a weighted network and consider it as our null theory, so it follows that $A_{ij} = C_{ij}$ which we insert into the modularity function to give

$$Q = \frac{1}{2m} \sum_{i,j} (C_{ij} - < C_{ij} >) \delta(\mu_i, \mu_j) \tag{41}$$

This, however, yields a problematic null modell if defined simply as in the unweighted case. For correlation matrices, a common but problematic null model derived from configuration model logic is:

$$< C_{ij} >_{null} = \mathrm{Corr}[X_i, X_{\mathrm{tot}}] \cdot \mathrm{Corr}[X_j, X_{\mathrm{tot}}]. \tag{42}$$

which is inconsistent with properties of correlations and hence we argue that we cannot directly treat correlation matrices as standard unweighted/weighted networks in the Newman-Girvan modularity sense without careful null model definition. Hence, results show that modularity is often not of proper use to detect communities from raw correlation matrices, and complex network clustering approaches need adaptation. To overcome this limitation we return to RMT to obtain improved and consistent methods to cluster many time series via a null hypothesis that fits appropriately.Then, what null hypothesis can we take? To create a set of modularity function which can be used to identify communities we redefine the modularity function making use of the results of the RMT. This application was carried out originally by Mel MacMahon1 and Diego Garlaschelli in their article [13]. We set the null model for the correlation matrix to be part explained by market-wide effects and noise, i.e., C_{null}:

$$< C_{ij} >_{null} = C_{ij}^{overall} + C_{ij}^{noise} \tag{43}$$

Note that the sectoral behavior has not been included in the null hypothesis as that is the behavior we are trying to obtain by negating the null hypothesis. Now we can then rewrite the modularity function to get [12]

$$Q^{(\ell)}(\sigma) = \frac{1}{Z} \sum_{i,j} \tilde{C}_{ij}^{(\ell)} \cdot \delta(\sigma_i, \sigma_j) \tag{44}$$

$$\tilde{C}^{(\ell)} = C^{(\ell)} - \langle C^{(\ell)} \rangle_{null} \tag{45}$$

The filtered matrices $C^{(\ell)}$ correspond to different parts of the spectrum:

$$C^{(\ell)} = \begin{cases} C^{\mathrm{market}} = C^{(1)} \\ C^{\mathrm{noise}} = C^{(2)} \\ C^{\mathrm{sector}} = C^{(3)} \end{cases} \tag{46}$$

where we refer to $\mathbf{C}^{(l)}$ as the filtered correlation matrix. Here we have set Z to be a normalization factor, often related to the sum of all weights or correlations in the part of the matrix being analyzed. The modularity is given by the sum of the intracommunity correlations (in $\tilde{C}^{(\ell)}$) and divided by X_{tot}, this last being a measure of the volatility of the system. Note also that more complex null models are possible by incorporating them in $< C_{ij} >_l$. Thus, the main takeaway point is that we can choose whatever null model we prefer as long as it obeys certain properties of correlation matrices, which the standard network-based definition of modularity does not inherently satisfy for correlation data.

5 Application of the model

With all the information presented we can make use of the Random Matrix Theory to filter the correlation matrix and then compute the modularity optimization task to obtain the community structure we have long talked about. This way, we would obtain stocks which are grouped together by their co-movement rather than by market modes or noise. [2]

We have, then, achieved a phenomenological theory to describe financial markets, rather than trying to take into account every single piece affecting the economical structure, which would inevitably fall under over-complexity. We have also built an alternative method to both Brownian motion and Levy flights to describe markets by analyzing not one market but many markets at the same time, which as we saw is inevitable as markets are deeply correlated.

But why is this all so important? What is the application of grouping stocks together by correlation? The straightforward answer lies on Modern Portfolio Theory [6]. This theory uses correlation matrices as a measure of risk estimation, which is key to optimizing gains and reducing losses in markets. A very brief introduction was made in the extended version of the paper, but further documentation should be found in the literature where this topic is fruitfully being treated.

For the markets introduced throughout this work we can develop a clustering of the form:

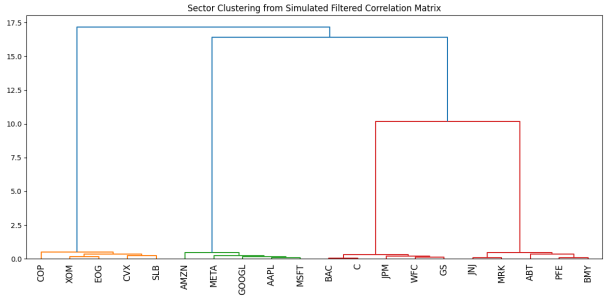

Figure 33: Community clustering from the correlation matrix obtained previously. Different sectors are clustered together

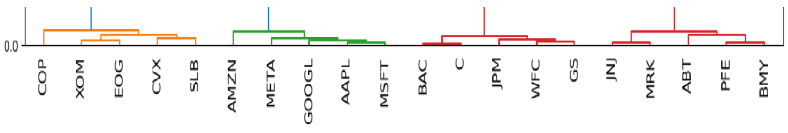

Figure 34: Zoomed view of the clusters in Figure 33.

Sector clustering is obvious, where finance markets, energy markets or else are grouped together, validating modularity optimization to cluster markets, along with RMT procedures to filter noise and create a sectoral behavior. These models can be used as an introduction to risk estimation in Modern Portfolio Theory!

6 Conclusions

This work represents a natural evolution of the current state-of-the-art in econophysics. Brownian motion was first introduced as the basic model and served as a baseline for the remainder of the study. Lévy flights continued the explanatory framework, but the need arose for more complex systems involving Random Matrix Theory (RMT). This investigation has made it possible to address the limitations of classical approximations, particularly their inability to capture extreme events, correlations among stocks, or memory effects—all of which are present in real financial markets.

It has become clear that financial markets exhibit a highly complex structure that often cannot be adequately modeled by the Efficient Market Hypothesis (EMH). Through Python simulations and empirical analysis, it has been observed that financial time series do not follow purely stochastic dynamics. Instead, correlations, memory, and other effects are evident. These findings challenge the validity of the EMH and justify the need for models that incorporate such features.

It has been highlighted that various tools from statistical physics are highly valuable for analyzing financial markets. In this context, econophysics presents itself as a promising framework for modeling agent interactions and detecting hidden patterns that reveal important market features. Analogies with nuclear systems and the application of spectral matrix theory offer a rich and realistic description of financial systems.

Although the aim of this work was not to present a new theoretical model, it provides a structured narrative of the intellectual evolution within the field of econophysics, revisiting foundational concepts and offering a thoughtful reassessment of current models. Custom Python simulations have supported the theoretical arguments and demonstrated the practical application of the concepts.

Ultimately, this work seeks to highlight the need to move beyond overly simplistic models in favor of multidimensional, empirically grounded approaches. Financial studies must evolve by integrating methodologies from disciplines as intricate as physics. This interdisciplinary effort would lead to more accurate representations of features that are of utmost importance in understanding financial markets.

References

[1] Juan M. Aguirregabiria Aguirre. *Fisika ikasleentzako ekuazio diferentzial arruntak*. Universidad del País Vasco / Euskal Herriko Unibertsitatea, Argitalpen Zerbitzua, Bilbo, 1ª ed., 1ª imp. edition, 2000. Unibertsitateko Eskuliburuak – Manuales Universitarios.

[2] Tomaso Aste, Tiziana Di Matteo, Michele Tumminello, and Rosario N. Mantegna. Correlation filtering in financial time series (invited paper). In D. Abbott, J.-P. Bouchaud, X. Gabaix, and L. McCauley, J. editors, *Noise and Fluctuations in Econophysics and Finance, Proceedings of SPIE, Vol. 5848*, pages 100–109, Bellingham, WA, 2005. SPIE – The International Society for Optical Engineering.

[3] Louis Bachelier. *Théorie de la spéculation*. Annales scientifiques de l'École normale supérieure, 3e série, tome XVII. Gauthier-Villars, Paris, 1900. Tesis de doctorado; sentó las bases del modelado estocástico en finanzas.

[4] Carlos A. Bertulani. *Nuclear Physics in a Nutshell*. Princeton University Press, Princeton, NJ, 2007.

[5] Aashish A. Clark. Quantum noise and quantum measurement. `URLofthePDF`, 2021. PDF; retrieved 13 December 2021.

[6] Carlo Requião da Cunha. *Introduction to Econophysics: Contemporary Approaches with Python Simulations*. CRC Press, Taylor & Francis Group, Boca Raton, FL, USA, 2022.

[7] Robert P. Dobrow. *Introduction to Stochastic Processes with R*. John Wiley & Sons, Inc., 2016. First published: 11 March 2016; Online ISBN: 978-1-118-74071-2.

[8] Eugene F. Fama. Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, 25(2):383–417, May 1970. Papers and Proceedings of the Twenty-Eighth Annual Meeting of the American Finance Association, New York, N.Y., December 28–30, 1969.

[9] Sofien Kaabar. *Deep Learning for Finance: Creating Machine & Deep Learning Models for Trading in Python*. O'Reilly Media, Inc., Sebastopol, CA, 2024. Released January 2024; code repository at `https://github.com/sofienkaabar/deep-learning-for-finance`.

[10] L. Laloux, P. Cizeau, J.-P. Bouchaud, and M. Potters. Noise dressing of financial correlation matrices. *Physical Review Letters*, 83(7):1467–1470, 1999.

[11] Andrew W. Lo. Long-term memory in stock market prices. *Econometrica*, 59(5):1279–1313, 1991.

[12] Thomas Lux and Michele Marchesi. Scaling and criticality in a stochastic multi-agent model of a financial market. *Nature*, 397(6719):498–500, 1999.

[13] Mel MacMahon and Diego Garlaschelli. Community detection for correlation matrices. *Physical Review X*, 5(2):021006, 2015.

[14] Rosario N. Mantegna and H. Eugene Stanley. Scaling behaviour in the dynamics of an economic index. *Nature*, 376:46–49, 1995.

[15] Naia Ormaza Zulueta and Josu Mirena Igartua. Prozesu estokastikoak fisikan eta ekonomian: higidura browndarretik finantza-merkatuen ereduetara. *Ekaia EHUko Zientzia eta Teknologia aldizkaria*, 39:257–276, 2021.

[16] Vasiliki Plerou, Parameswaran Gopikrishnan, Bernd Rosenow, Luís A. Nunes Amaral, Thomas Guhr, and H. Eugene Stanley. Random matrix approach to cross correlations in financial data. *Physical Review E*, 65(6):066126, 2002. Published 27 June 2002.

[17] Sitabhra Sinha, Arnab Chatterjee, Anirban Chakraborti, and Bikas K. Chakrabarti. *Econophysics: An Introduction*. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 2010.

[18] Petre Stoica and Randolph L. Moses. *Spectral Analysis of Signals*. Pearson/Prentice Hall, Upper Saddle River, NJ, 2005. Includes bibliographical references and index.

7 SDG

This work provides both theoretical and computational tools aimed at addressing complex topics within the field of econophysics, making them more accessible and comprehensive. In doing so, it contributes to Sustainable Development Goal (SDG) 8 by enhancing the understanding of financial markets. Accurate models based on the discussions in this work can foster more sustainable economic growth. Furthermore, by questioning outdated approaches and presenting up-to-date insights from the field, it opens the door to more informed economic policies and financial decision-making.

The physical approach adopted in this study contributes to scientific and technological innovation in the field of economics. It aligns with SDG 9 by promoting interdisciplinary research and encouraging the adoption of more robust and accurate models in crucial areas such as risk management.

The analysis of financial returns and the presence of extreme events (fat tails) highlights how vulnerable groups can be disproportionately affected by traditional financial models, which often underestimate risks with significant social impact. A deeper understanding of financial dynamics can support the development of policies aimed at mitigating the effects of financial crises on disadvantaged sectors, thereby contributing to the reduction of inequalities. All of these fall under the premise of SDG 10 (reduced inequalities).

SDG 17 is clearly reflected in this work, which adopts an interdisciplinary approach that integrates knowledge from physics, mathematics, and economics. Collaboration across these fields is essential to achieving meaningful results, exemplifying the spirit of SDG 17, which promotes partnerships among academic institutions. By encouraging an innovative and integrative approach to financial analysis, this work lays the foundation for cooperative efforts that can support sustainable development and foster more resilient economic systems.

Figure 35: The 17 Sustainable Development Goals adopted by the UN in the 2030 Agenda.