AIMS Mathematics, 9(12): 34947–34965.
DOI:10.3934/math.20241663
Received: 30 September 2024
Revised: 24 November 2024
Accepted: 02 December 2024
Published: 16 December 2024

https://www.aimspress.com/journal/Math

*Research article*

# Stock volatility as an anomalous diffusion process

**Rubén V. Arévalo[1], J. Alberto Conejero[1], Òscar Garibo-i-Orts[1,2] and Alfred Peris[1,*]**

[1] I.U. de Matemàtica Pura i Aplicada, Universitat Politècnica de València, 46022 Valencia, Spain; rubenvicentearevalo@gmail.com, aconejero@upv.es, aperis@mat.upv.es

[2] GRID—Grupo de Investigación en Ciencia de Datos, Valencian International University—VIU, 46002 Valencia, Spain; oscar.garibo@professor.universidadviu.com

* **Correspondence:** Email: aperis@mat.upv.es.

**Abstract:** Anomalous diffusion (AD) describes transport phenomena where the mean-square displacement (MSD) of a particle does not scale linearly with time, deviating from classical diffusion. This behavior, often linked to non-equilibrium phenomena, sheds light on the underlying mechanisms in various systems, including biological and financial domains.

Integrating insights from anomalous diffusion into financial analysis could significantly improve our understanding of market behaviors, similar to their impacts on biological systems. In financial markets, accurately estimating asset volatility—whether historical or implied—is vital for investors.

We introduce a novel methodology to estimate the volatility of stocks and similar assets, combining anomalous diffusion principles with machine learning. Our architecture combines convolutional and recurrent neural networks (bidirectional long short-term memory units). Our model computes the diffusion exponent of a financial time series to measure its volatility and it categorizes market movements into five diffusion models: annealed transit time motion (ATTM), continuous time random walk (CTRW), fractional Brownian motion (FBM), Lévy walk (LW), and scaled Brownian motion (SBM).

Our findings suggest that the diffusion exponent derived from anomalous diffusion processes provides insightful and novel perspectives on stock market volatility. By differentiating between subdiffusion, superdiffusion, and normal diffusion, our methodology offers a more nuanced understanding of market dynamics than traditional volatility metrics.

**Keywords:** volatility; anomalous diffusion; recurrent neural network; stock markets; AnDi-Challenge
**Mathematics Subject Classification:** 91G80, 60J60, 68T07

## 1. Introduction

Understanding asset volatility is paramount in investment decisions, ranking alongside expected returns as a fundamental parameter. Related to the returns forecasts and simulation, from Charles Dow's late 19th-century theory of three market trends [1] to contemporary financial time series forecasting using deep learning [2–5], market participants have long endeavored to predict financial market prices. However, the vast amount of information and the myriad of influencing variables render most models inaccurate in predicting asset price movements. Moreover, any published model capable of forecasting asset prices would likely see its predictive power diminish as market participants begin to exploit it, thus altering the market dynamics upon which it was trained [6, 7].

Currently, stochastic processes are the most widely accepted models for simulating the price behavior of financial assets [8]. These stochastic models are founded on the Brownian motion, first observed in 1827 by Robert Brown, while examining pollen particles from *Clarkia pulchella* through a microscope [9]. The mathematical description of Brownian motion through the stochastic differential equation for the position $x(t)$ of a particle over time $t$ is given by:

$$dx(t) = \mu dt + \sigma dW(t),$$

where $x(t)$ represents the particle position at time $t$; $\mu$ is the drift coefficient, indicating the central tendency of the motion; $\sigma$ is the diffusion coefficient, reflecting the volatility of the motion; and $dW(t)$ signifies the increment of a Wiener process (or Brownian motion), which represents the stochastic component of the motion.

The erratic movement of a pollen particle, colliding with a myriad of smaller particles moving at various velocities in random directions, exemplifies Brownian motion. The complex interactions underlying this pattern cannot be fully resolved by models accounting for each molecule involved. Thus, only probabilistic models considering entire molecular populations are viable for describing this phenomenon, as was first conjectured by Einstein [10] and later verified by Perrin [11].

This same phenomenon occurs in financial markets, where a myriad of variables with varying influences impact asset prices. Therefore, advancements in the field of stochastic processes have direct applications in financial markets. The most notable example is the Black-Scholes-Merton model (BSM), which is derived from Brownian motion and assumes that the stock price follows a geometric Brownian motion [12]. It is widely used in financial mathematics to model the dynamics of financial instruments, especially for European option pricing [13].

The BSM model revolutionized financial markets by providing a mathematical framework for pricing European-style options, particularly call options. A call option gives the holder the right, but not the obligation, to buy a specific asset (often a stock) at a predetermined price (strike price, $K$) on a specified expiration date $T$. The need to price these options arises because they are traded in financial markets, and their value fluctuates with changes in the underlying asset's price $S(t)$, time to expiration $t$, volatility $\sigma$, and interest rates $r$.

Pricing options during their life is crucial for investors, traders, and financial institutions as it helps manage risk, structure portfolios, and make informed trading decisions. The call option price must reflect the current market conditions, expectations, and the probability that the option will be profitable at its maturity, i.e., the asset's price at expiration will be above the strike price.

At maturity, the formula for a European call option's value depends only on the relationship between

the underlying asset's price and the strike price. Specifically, the call option will only have value if the underlying asset's price exceeds the strike price. Otherwise, the option expires worthless. The payoff at maturity for a call option is given by:

$$C(T) = \max(S(T) - K, 0). \tag{1.1}$$

The BSM formula for the price of a call option for a time previous to its expiration is:

$$C(S, t) = S(t)N(d_1) - Ke^{-r(T-t)}N(d_2), \tag{1.2}$$

where:

- $N(d_1)$ and $N(d_2)$ are the cumulative distribution functions of the standard normal distribution evaluated at $d_1$ and $d_2$, which reflect the probability of the option being in-the-money (ITM, profitable at expiration),
- $e^{-r(T-t)}$ is the discount factor, which adjusts the strike price for the time value of money, since a payoff in the future is worth less than an equivalent payoff today,
- $d_1$ represents the logarithmic difference between the current asset price and the strike price, adjusted for the risk-free rate and volatility over the time to expiration. Intuitively, this can be interpreted as a measure of how far the current price is from the strike price in terms of standard deviations, considering the time remaining until expiration and the volatility of the asset,

$$d_1 = \frac{\ln(S/K) + (r + \sigma^2/2)(T - t)}{\sigma\sqrt{T - t}} \tag{1.3}$$

- $d_2$ is $d_1$ adjusted by the volatility term. It represents the probability-adjusted distance between the current asset price and the strike price, accounting for the uncertainty (volatility) of future price movements over the time until the option expires. In essence, this corresponds to the probability that the option will expire in the money under the risk-neutral measure (the standard assumption in option pricing models).

$$d_2 = d_1 - \sigma\sqrt{T - t} \tag{1.4}$$

These expressions capture the probability-adjusted price dynamics of the underlying asset, incorporating its expected growth and the uncertainty around its future value (i.e., volatility).

While the BSM model assumes that volatility is a known constant, in practice, implied volatility (IV) is derived from the market prices of options. Implied volatility reflects the market's expectation of future volatility and is a critical input for pricing options. It is "implied" because it is inferred from the option's current price rather than directly observed. Originating from the BSM, there are two primary methods to derive the implied volatility, which reflects the market's expectation of volatility based on stock prices:

1. *Numerical methods*: These are critical because implied volatility is not directly observable in the market and must be inferred from the prices of traded options. These numerical techniques iteratively adjust the volatility input until the model and market prices align. For instance, these have been explored by Deng, as detailed in [14] and Nabubie and Song in [15].

2. *Closed-form approximations* refer to an analytical expression that provides an explicit formula for calculating the volatility, as opposed to requiring iterative numerical methods. There have been several studies in this area, such as [16] and [17], and further developed in more recent studies, such as those referenced in [18].

Currently, stochastic models have extended beyond option pricing and modernized the analysis across various asset types [13]. Moreover, in recent years, various studies have utilized new machine learning (ML) [19, 20] and deep learning techniques [21, 22] to predict volatility.

For a stock whose price follows a Wiener process, the width of the distribution increases as time increases. The probability of the asset's price deviating farther from its initial value increases over time. However, rather than increase linearly, the volatility increases with the square root of time. Since observed price changes do not always follow Gaussian distributions [23], others, such as Lévy distributions, are often used [24], which can capture attributes such as *fat tails*.

The formulas used in calculating the actual historical volatility, explained in Section 3, are accurate extrapolations of a random walk (a Wiener process) whose steps have finite variance. However, more generally, for natural stochastic processes, the precise relationship between volatility measures for different time periods is more complicated. Some use the Lévy stability exponent $\alpha$ to extrapolate natural processes:

$$\sigma_T = T^{1/\alpha}\sigma,$$

where:

- $\sigma_T$ is the volatility or standard deviation of the process over a time period $T$. It is the measure of how much the process fluctuates over a given period,
- $T$ is the time period over which volatility is being measured,
- $\alpha$ is the Lévy stability exponent or scaling exponent. It characterizes the type of stochastic process and the "stability" of the distribution of the process. For a normal random walk, $\alpha = 2$, which leads to the standard relationship between volatility and time, where volatility scales with the square root of time:
  - If $\alpha = 2$, the process behaves like a Brownian motion, with normal diffusion.
  - If $0 < \alpha < 2$, the process follows a Lévy flight, and the volatility scales differently due to "fat tails" in the distribution of the steps.
  - A smaller $\alpha$ leads to more significant jumps in the process, reflecting the possibility of extreme events (non-Gaussian behavior).
- $\sigma$ is the volatility over a unit time period.

Some people believe $\alpha < 2$ for financial activities such as stocks, indexes, etc. Mandelbrot discovered this when he looked at cotton prices and found that they followed a Lévy alpha-stable distribution with $\alpha = 1.7$ [25].

Ultimately, these volatility measures aim to quantify how much the prices of stocks vary, but this measure is also necessary in many other fields, such as biology when studying the movement of a cell [26], mutations [27, 28], or transport in porous media [29]. So, it is of interest to apply the models developed in other fields to finance.

Some of these developments have been the models developed for the AnDi Challenge to characterize anomalous diffusion processes through machine learning and statistical methods [30, 31].

Anomalous diffusion is a prevalent phenomenon linked to non-equilibrium states, including energy and information flows [32, 33]. In a similar way, skewness and kurtosis may characterize the changing properties of a system [34–36].

To describe diffusion, we can consider the following analogy: Let us imagine a particle being an ant or some other diminutive explorer; we can then think of MSD, which can be written as $\langle x^2 \rangle$, as the portion of the system that it has explored. For normal diffusion such as Brownian motion, the relationship between the portion of the explored region and time is linear, $\langle x^2 \rangle \sim t$. As time progresses, the expected value of distance explored by our ant (MSD) will remain constant. In contrast to normal diffusion, anomalous diffusion is characterized by $\langle x^2 \rangle \sim t^\alpha, \alpha \neq 1$, where $\alpha$ is the anomalous diffusion exponent, or 2 times the Hurst exponent. Anomalous diffusion can be further subdivided into superdiffusion and subdiffusion when $\alpha > 1$ or $\alpha < 1$, respectively [37, Ch. 6].

In this work, we first connect the terminologies of the $\beta$ coefficient used when estimating asset and market volatilities and the exponent of anomalous diffusion; see Section 2. This connection lets us analyze market time series as if they were particles moving in space, whose exponent (volatility) can be estimated and whose underlying diffusion model can be estimated. The machine learning methods used for the exponent regression and classification tasks will be described in Section 4. These methods will be used on the data that is described in Section 3,

In Section 5 we will examine the exponent evolution during various events and across different industries. Additionally we explore the correlation between the exponent and both historical and implied volatility. Also, identifying the model behind the trajectory is crucial, as it provides substantial information on system dynamics; this analysis will be conducted in Section 5.1.2. We end with conclusions in Section 6.

## 2. Volatility in financial markets

Stock price evolution in financial markets can be considered as a non-equilibrium steady state. On the one hand, the constant information flux from economic, political, and social events leads to price adjustments and trading activities, which avoid that it would lead to an equilibrium. On the other hand, volatility periods exhibit memory effects and non-Markovian dynamics. This was already known when studying the Hurst exponent for showing the presence of long-memory effects [38–40]. Besides, an analysis of the presence of anomalous diffusion (subdiffusion and superdiffusion) is consistent with non-equilibrium dynamics.

In an anomalous diffusion process, the MSD grows following a power law of the form:

$$y(t) = at^\gamma, \tag{2.1}$$

where $a$ is the proportionality constant, and $\gamma$ is the power exponent. If we take a logarithmic transformation, this can be interpreted as a linear regression:

$$Y(t) = \log(y(t)) = \log(a) + \gamma \log(t) = c + \gamma X(t) \tag{2.2}$$

where $Y(t) = \log(y(t))$, $c = \log(a)$, and $X(t) = \log(t)$. This equation follows parallelism with the returns

of the asset $i$ at time $t$ which are given by the relationship:

$$r_{i,t} = \alpha_i + \beta_i \cdot r_{m,t} + \varepsilon_t \tag{2.3}$$

where $r_{i,t}$ is the rate of return of an asset $i$ at time $t$, $r_{m,t}$ is the rate of return of the stock-market index $m$ at time $t$, and $\varepsilon_t$ is an unbiased error term whose squared error should be minimized. The coefficients $\beta_i$ and $\alpha_i$ are usually referred to as the $\beta$ and $\alpha$ coefficients. Applying the ordinary least squares fitting, we obtain that the solution for the beta coefficient is:

$$\beta_i = \frac{\mathrm{Cov}(r_i, r_m)}{\mathrm{Var}(r_m)}. \tag{2.4}$$

This concept of the stock beta coefficient is rooted in the capital asset pricing model (CAPM) [41], a model that defines the expected return from a portfolio of stocks with varying degrees of risk. It also considers the volatility of a particular security in relationship to the market.

Understanding the beta coefficient is crucial for investors looking to assess the risk associated with a particular stock compared to the broader market. Let us consider the time series of market $r_m = \{r_{m,1}, \ldots, r_{m,t}\}$ and asset $r_i = \{r_{i,1}, \ldots, r_{i,t}\}$ returns. Therefore, by applying the definitions of standard deviation, variance, and correlation, we compute the beta coefficient as follows:

$$\beta_i = \rho_{i,m} \frac{\sigma_i}{\sigma_m},$$

where $\sigma_i and \sigma_m$ are the standard deviations of the asset and market time series returns, also known as *volatilities*, and $\rho_{i,m}$ is the correlation of the two market and asset returns that is computed as

$$\rho_{i,m} = \frac{\mathrm{Cov}(r_i, r_m)}{\sqrt{\mathrm{Var}(r_i)\,\mathrm{Var}(r_m)}}.$$

Therefore, the beta coefficient is a key component in selecting securities for a diversified portfolio, allowing investors to balance potential returns against their risk tolerance. The first thing to notice about the beta coefficient of a stock $i$, $\beta_i$, is whether it is less than or greater than one:

- If $\beta_i > 1$, it indicates that a stock is more volatile than the market, offering the potential for higher returns but with greater risk, as with more cyclical companies.
- If $\beta_i = 1$, it indicates that it has the same risk as the market.
- If $0 < \beta_i < 1$, it suggests that the stock is less volatile than the market, potentially offering more stable returns but with less risk, as with non-cyclical companies.
- If $\beta_i = 0$, it suggests no correlation with market movements.
- If $\beta_i < 0$, it indicates an inverse relationship with the market.

This relationship can also be observed in the exponent of the anomalous diffusion process since exponents greater than one define superdiffusion processes and exponents less than one define subdiffusion processes. Therefore, unlike conventional volatility, estimating the volatility calculated through the anomalous diffusion exponent will allow us to know what kind of process our stock follows.

Although both $\beta$ and $\alpha$ measure aspects of volatility, they represent different perspectives on the system. $\beta$ measures relative risk, reflecting how a stock moves to the whole market. In contrast, the diffusion exponent characterizes the nature of volatility over time, indicating whether the stock follows a subdiffusive, normal-diffusive, or superdiffusive process. While $\beta$ assesses comparative risk, $\alpha$ provides insights into the underlying dynamics of market movements.

In the context of stock markets, an intuitive example of superdiffusion would be when a stock enters into a bubble or the economy is at its peak, increasing its volatility. Thus, the relationship of returns and time will be $\langle x^2 \rangle \sim t^\alpha, \alpha > 1$. Conversely, if the stock arrives at a stable period, showing a decrease in its volatility, it will follow a subdiffusion process [42].

This connection not only highlights the ability of the anomalous diffusion framework to characterize market behaviors but also allows us to relate these findings to established measures of persistence and memory in financial time series, such as the Hurst exponent.

The Hurst exponent ($H$) is directly related to the anomalous diffusion exponent ($\alpha$), with $H = \alpha/2$. For Brownian motion, where $\alpha = 1$, the Hurst exponent equals $H = 1/2$, representing normal diffusion. This relationship bridges financial time series' long memory analysis with anomalous diffusion processes [43].

## 3. Historical data

For our analysis, we selected the closing prices for companies listed in the S&P 500, owing to its recognition as one of the most closely monitored indexes worldwide. This choice guarantees higher data accuracy and reliability compared to other indexes. Furthermore, we categorized these companies* by their respective industries, which will facilitate a more detailed analysis in subsequent sections.

**Table 1.** S&P 500 data charaacteristics.

| | |
|---|---|
| Total number of companies | 484 |
| Industries | 11 |
| First historical day | 2012-03-29 |
| Last historical day | 2024-03-01 |

Once we have the historical data, we must compute the returns from our starting stock price. We start at 0, and thus reflect the cumulative returns at each step. With the normalization of the data, we could use the models from the AnDi Challenge, described in Section 4, as they were trained in trajectories starting at zero with the size jumps at each step.

Those returns have been computed for windows between [10–20], [100–200], [400–500], and [800–1000] days, over intervals of four years, shifting one month at a time. This approach enabled us to compute the returns for 95 intervals to see the exponent variation over time.

It is important to note that not all companies have data spanning the entire historical period under review, concluding with a dataset of 171,793 time series or trajectories in the terminology of diffusion processes. To compare our estimated stock volatility using the anomalous diffusion exponent with the market volatility, we must first obtain the historical and implied volatility.

We recall that volatility, already introduced as $\sigma$, is the degree of price variation of an asset over

*There are not 500 companies because we do not have enough historical closing prices for the 16 missing companies.

time, usually measured by the standard deviation of logarithmic returns. Historic volatility measures a time series of past market prices, whereas implied volatility looks forward in time, being derived from the market price of a market-traded derivative, as explained next. In particular, volatility could refer to different concepts:

- *Actual current volatility* refers to the standard deviation of the returns for a specified historical period, with the last observation being today.

- *Actual historical volatility* refers to the standard deviation of the returns of an asset over a specified period. This will be the *realized volatility*, and this is the one we will use in our analysis. The general formula is:

$$\sigma_T = \sigma_{annually} \sqrt{T}, \tag{3.1}$$

  where $\sigma$ is the square root of the realized variance as mentioned in Equation 3.1 and $T$ the time horizon. Therefore, the annualized volatility is:

$$\sigma_{annually} = \sigma_{daily} \sqrt{252}, \tag{3.2}$$

  where 252 are the trading days in a given year. This calculation has been applied to the datasets described in Table 2.

**Table 2.** Dataset specifications.

| | |
|---|---|
| Total number of companies | 484 |
| Periods | 95 |
| Different window sizes | 4 |
| **Total dataset** | **183,920** |

- Actual future volatility or implied volatility refers to the volatility of an asset over a specified period starting at the current time and ending at a future date.

  – Historical implied volatility refers to the implied volatility observed from historical prices of the options, which will be the other volatility we will use to compare our exponent.

  – Current implied volatility refers to the implied volatility observed at the current date. An example of this is show in Figure 3. The headers in Table 3 with the percentage and the index price are the call moneyness and the strikes of the options, i.e. $K$ in Eqs 1.2, 1.3, and 1.4. Moneyness represents the percentage variation of the strike price from the current market price, being 100% of the at-the-money (ATM) price. The index column represents the maturity of the options, i.e., $T$.

**Table 3.** Example of the implied volatility for the OMX index.

| **Moneyness** | **80%** | **90%** | **95%** | **97.5%** | **100%** | **102.5%** | **105%** | **110%** | **120%** |
|---|---|---|---|---|---|---|---|---|---|
| **Index Price** | **1451** | **1633** | **1724** | **1769** | **1800** | **1860** | **1905** | **1996** | **2177** |
| **23 Oct 2020** | 36.54 | 36.71 | 33.09 | 29.40 | 25.41 | 22.26 | 22.56 | 23.70 | 27.10 |
| **29 Oct 2020** | 33.34 | 32.47 | 28.16 | 24.98 | 21.71 | 19.00 | 18.78 | 19.75 | 22.53 |
| **19 Mar 2021** | 30.73 | 25.92 | 23.64 | 21.78 | 19.77 | 21.43 | 21.84 | 20.45 | 22.50 |
| **18 Jun 2021** | 28.23 | 24.27 | 22.31 | 21.33 | 20.37 | 19.44 | 18.58 | 17.15 | 19.90 |
| **17 Dec 2021** | 26.01 | 22.73 | 21.18 | 20.44 | 19.72 | 19.02 | 18.36 | 17.15 | 15.38 |
| **16 Dec 2022** | 23.66 | 21.53 | 20.53 | 20.05 | 19.60 | 19.17 | 18.77 | 18.07 | 17.08 |
| **15 Dec 2023** | 22.37 | 20.72 | 19.94 | 19.58 | 19.23 | 18.91 | 18.61 | 18.11 | 17.45 |
| **20 Dec 2024** | 21.63 | 20.20 | 19.54 | 19.24 | 18.95 | 18.68 | 18.44 | 18.02 | 17.48 |
| **30 Dec 2024** | 21.62 | 20.20 | 19.55 | 19.24 | 18.95 | 18.69 | 18.44 | 18.03 | 17.48 |
| **30 Dec 2025** | 21.13 | 19.84 | 19.27 | 19.00 | 18.75 | 18.52 | 18.31 | 17.94 | 17.43 |
| **30 Dec 2026** | 20.85 | 19.65 | 19.12 | 18.88 | 18.65 | 18.44 | 18.24 | 17.90 | 17.40 |
| **30 Dec 2027** | 20.62 | 19.50 | 19.01 | 18.78 | 18.57 | 18.37 | 18.19 | 17.86 | 17.35 |
| **29 Dec 2028** | 20.43 | 19.36 | 18.90 | 18.69 | 18.49 | 18.30 | 18.12 | 17.80 | 17.28 |
| **28 Dec 2029** | 20.26 | 19.24 | 18.80 | 18.60 | 18.40 | 18.22 | 18.05 | 17.73 | 17.21 |

Finally, we proceed to obtain the historical implied volatility for the 484 stocks with a strike price of 100% and maturity equal to 30 days.

## 4. Methods

Identifying the underlying mechanisms and classifying the processes involved are critical steps in understanding anomalous diffusion. However, measuring these properties through trajectory data analysis poses significant challenges, especially when dealing with short, irregularly sampled, or mixed behavior trajectories. Recently, several methods have been proposed to quantify the anomalous diffusion process beyond the classical mean-square displacement calculation [44–51].

In this work, we have chosen one of the top-performing models in the AnDi Challenge [45] to estimate the anomalous diffusion exponent of the MSD as a measure of volatility and to classify the type of diffusion process observed in the historical prices of companies listed on the S&P 500. The architecture used for both classification and regression tasks consists of a combination of convolutional and recurrent neural networks described in [45] and successfully applied in [52].

We briefly outline the model. First, two convolutional layers extract spatial features from the raw trajectories. An initial convolutional layer with 32 filters and a kernel size of 5, making a sliding window of size 5 through each trajectory, extracts spatial features from them. A second convolutional layer with more filters increased to 64 extracts higher-level features. We feed the resulting encoded trajectories into three consecutive bidirectional LSTMs to learn the sequential information, with a drop-out layer of 10% of the nodes to avoid incurring overfitting. Finally, we use several fully connected dense layers to predict the desired information (exponent regression or model classification).

For the regression method, we have used different trajectory lengths for the returns, with windows between [10–20], [100–200], [400–500], and [800–1000] days, which makes a total of 4 different

models, all sharing the same architecture. The model has been trained with the different window sizes and with $9 \cdot 10^6$ trajectories for the [10–20] model, $1.8 \cdot 10^6$ for the [100–200] model, and $9 \cdot 10^5$ for the [400–500] and [800–1000] models. These models were validated with $10^6$, $2 \cdot 10^5$, and $10^5$ trajectories, respectively. These split percentages were similar to the ones in [45]. All these trajectories were generated with the AnDi Dataset Library [53]. In the case of classification, we used a single model for all possible trajectory lengths and applied lead padding to each trajectory to make them of the same length (1000). This model was trained with $9 \cdot 10^5$ trajectories and validated with $10^5$ trajectories, too. White noise was added to the signal for all these cases with a signal-to-noise ratio equal to 2.

## 5. Results

To the best of our knowledge, this is the first time that the anomalous diffusion exponent $\alpha$ has been used for estimating stock price volatility. With the model explained in Section 4 and the historical returns obtained in Section 3, we can estimate each stock's volatility and different periods.

### *5.1. Results*

Once we have the evolution of $\alpha$ for each company and period, we can perform various analyses, such as the impact of certain events on the stock market and the exponents for different industries. To make the analyses more consistent, we have represented the evolution of the average of the companies by sector rather than do it for every single one of the 484 companies.

#### 5.1.1. Regression of the exponent $\alpha$

Figure 1 shows how more stable industries with less risk, such as Consumer Staples, have an exponent $\alpha < 1$. In contrast, more volatile sectors like energy show an exponent $\alpha > 1$, increasing in recent years. Additionally, we have represented with a black horizontal line the case of $\alpha = 1$ to delineate the processes of subdiffusion and superdiffusion.

On the other hand, it is noted that there are certain dates where there is a general increase in the exponent. Some of these dates are:

1. End of 2018: At the beginning of 2018, the main indices reached historical highs (this can be seen in the drop of the exponent at the beginning of 2018), but the trade war between the United States and China, along with the withdrawal of liquidity from central banks, caused drops in the markets and an increase in volatility.

2. Beginning of 2020: The COVID-19 crisis produced one of the largest market crashes at a speed never seen before.

Now, we will dive deeper into the correlation between the anomalous diffusion exponent $\alpha$ and two standard types of volatility: *actual historical volatility* (historical volatility) and *historical implied volatility* (implied volatility). So, we analyze the historical correlation between both volatilities and the exponent $\alpha$.

In Figure 2, we see that the historical volatility is very strongly correlated with the implied volatility, i.e., the historical volatility is aligned with the market-expected volatility. However, we can also see how historical and implied volatility correlate very highly with the $\alpha$ exponent. Moreover, to have

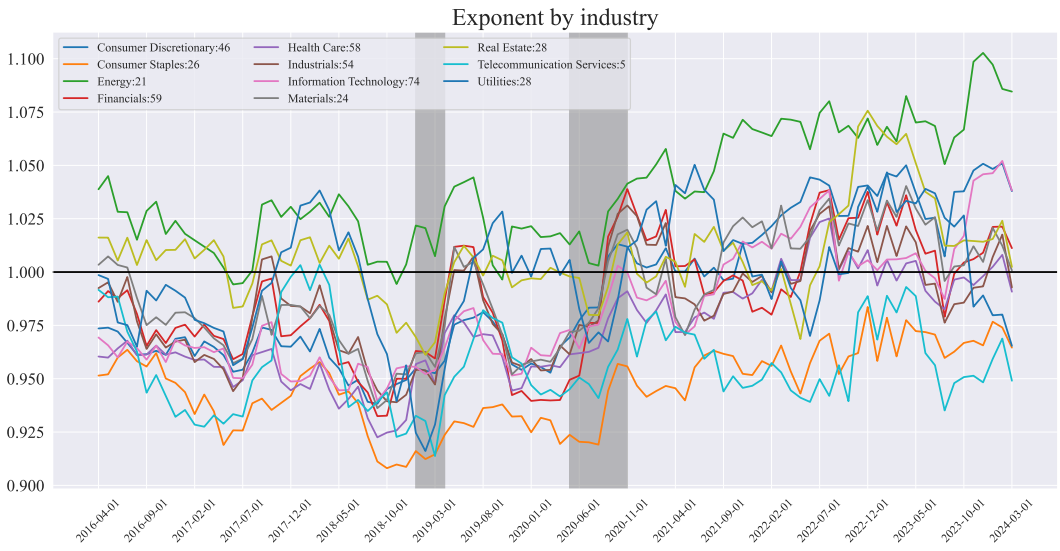


**Figure 1.** Exponent by sector and the number of companies in each one.

a more robust result, we have used the historical average to avoid having a correlation that is not consistent over time.

As shown in the top left subplot of Figure 2, the p-values are lower than 5%, indicating that the correlation is statistically significant. All three measures—historical volatility, implied volatility, and the $\alpha$ exponent—show a positive skew. This skewness is a common feature of stock returns, reflecting that while extreme positive returns are less frequent than smaller returns, they tend to be more probable than extreme negative returns. This asymmetric distribution is driven by investor behavior, sudden positive market shocks, or speculative bubbles that can cause sharp upward price movements. Research has extensively studied this characteristic of stock returns, as in [54]. These authors have examined the role of skewness in asset pricing and found that stocks with higher skewness tend to have higher expected returns. This indicates how the market prices the risk associated with skewed return distributions.

#### 5.1.2. Model prediction

Finally, we have classified the historical returns of the stocks according to the 5 models described in [30]. Since each underlying model has different properties, it is interesting to study to which model the stocks' historical returns align with. The five types of models are the following:

1. Annealed transit time motion (ATTM): This is a subdiffusive model with random variations in the diffusion coefficients along the time [55]. A particle, or stock price, diffuses with a diffusion coefficient $D_1$ during $T_1$ time units, and then it diffuses with a new diffusion coefficient $D_2$ during $T_2$ time units, and so on.

2. Fractional Brownian motion (FBM): This model covers the whole range of the $\alpha$ exponent (subdiffusive and superdiffusive). Motion is driven by a non-white (fractional Gaussian) noise,

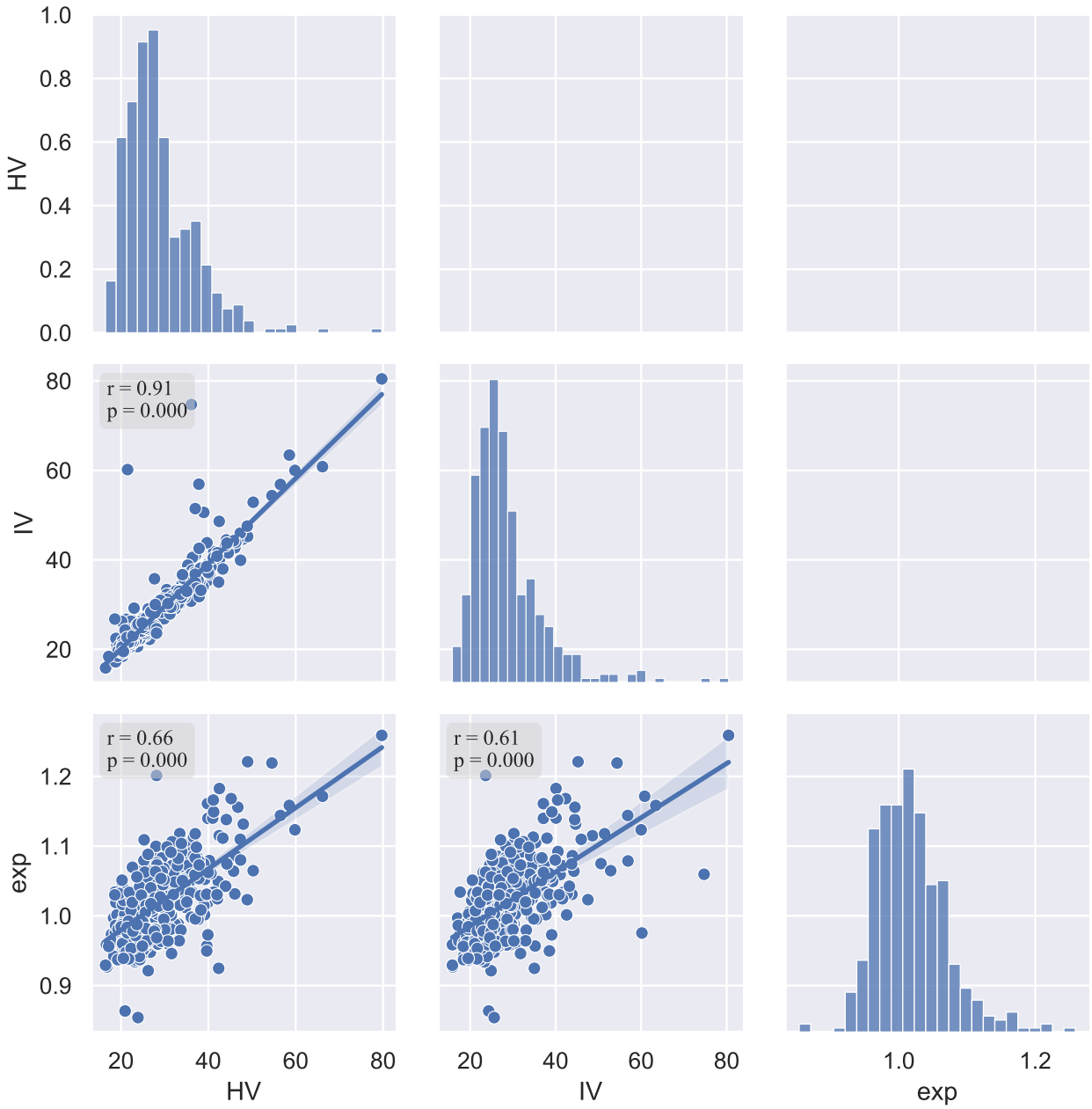


**Figure 2.** In the diagonal, a histogram distribution of the $\alpha$ exponent (exp) and the historical (HV) and implied volatility (IV). Below the diagonal, a scatter plot of these three measures and the correlation between them.

with a normal distribution with zero mean but power-law correlations between the noise at different times [43].

3. Continuous time random walk (CTRW): This subdiffusive model combines trap periods with no movement and a wide distribution of waiting times. Waiting times between displacements are sampled from a power-law distribution and displacements from a Gaussian probability distribution function with variance $D$ (the diffusion coefficient) and zero mean [56].

4. Lévy walk (LW): This superdiffusive model displays irregular jumps, correlated with the flight times, and with constant speed [57].

5. Scaled Brownian motion (SBM): This is also a subdiffusive and superdiffusive model whose diffusion coefficients depend on the time-related power-law exponent, and show a power-law dependence with it, and change constantly [55].

We classify the historical S&P 500 returns time series according to the model described in Section 4 into the aforementioned five diffusion types (ATTM, CTRW, FBM, LW, and SBM). By using these five models we intend to check if there are stocks whose price evolution in time differs from what FBM or BSM predict, offering new alternatives. For example, with ATTM and SBM, the diffusion exponent (or volatility) can vary over time. In the case of CTRW, we would have waiting times between price changes, in contrast to LW where we have more abrupt price changes.

All these changes happen conditioned by external factors. Since subdiffusion and superdiffusion processes can be associated with periods of lower and higher market volatility, respectively, in stock prices, the other two processes, FBM and SBM, which encompass both subdiffusion and superdiffusion behaviors, are the next most likely models. However, processes that solely describe superdiffusion (LW) or subdiffusion (CTRW) are the least likely. This is because the stock prices analyzed reflect periods of both high (superdiffusion) and low (subdiffusion) volatility. Consequently, a model that captures only one of these behaviors will fail to represent the full range of possible exponents.

We recall that we set 4 possible time windows of increasing length for the returns, [10–20], [100–200], [400–500], and [800–1000] days, over intervals of four years, shifting one month at a time, that cover from a few of weeks to some years. The resulting classification probabilities of each company are shown in Figure 3, and in Table 4, we can see how the results are consistent for each window size.

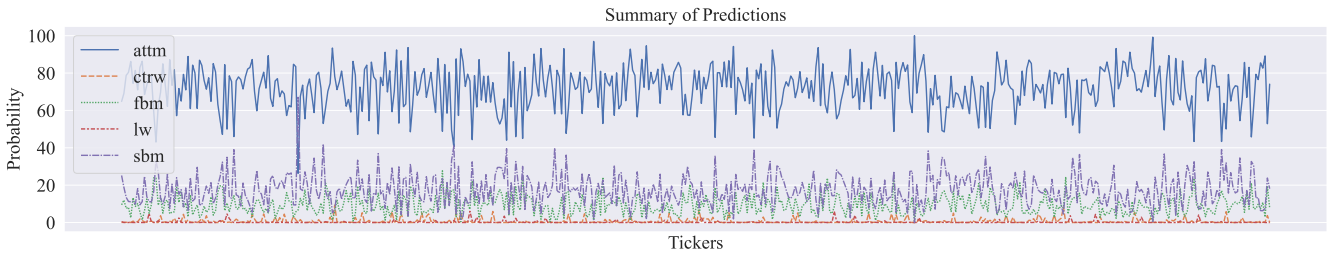


**Figure 3.** Classification of the S&P 500 stocks prices into one of the following models: ATTM, CTRW, FBM, LW, and SBM, depending on the length.

**Table 4.** Probabilities of each model classification regarding the window length.

| Window size | ATTM | CTRW | FBM | LW | SBM |
|---|---|---|---|---|---|
| 20 | 0.26 | 0.11 | 0.25 | 0.17 | 0.22 |
| 200 | 0.72 | 0.01 | 0.09 | 0.00 | 0.18 |
| 500 | 0.91 | 0.00 | 0.02 | 0.00 | 0.08 |
| 1000 | 0.82 | 0.00 | 0.01 | 0.00 | 0.18 |

The ATTM model explains the diffusion of particles in heterogeneous environments where the diffusion coefficient varies randomly over time. This means that instead of a constant diffusion rate, the system experiences changes in how quickly or slowly it diffuses based on different regions of time or space. In this case, these variations in diffusion coefficients mimic the behavior of stock returns over time, where the volatility of stock prices (analogous to the diffusion rate) can change unexpectedly due to market conditions, news, or investor behavior.

In the context of the S&P 500 returns, at any given time, certain stocks may experience periods of high volatility (large diffusion coefficients) or low volatility (small diffusion coefficients). Over longer time periods, the ATTM captures these random changes more effectively, so the model becomes more likely as the time window increases, see Table 4. Therefore, the ATTM is particularly useful for modeling financial markets because it allows for the dynamic nature of volatility, especially over extended periods.

Since subdiffusion and superdiffusion processes can be associated with periods of lower and higher market volatility, respectively, in stock prices, the other two processes, FBM and SBM, which encompass both subdiffusion and superdiffusion behaviors, are the next most likely models. However, processes that solely describe superdiffusion (LW) or subdiffusion (CTRW) are the least likely. This is because the stock prices analyzed reflect high (superdiffusion) periods and low (sub-diffusion) volatility. Consequently, a model that captures only one of these behaviors will fail to represent the full range of possible exponents.

## 6. Conclusions

In our research, we have observed a clear relationship between market historical volatility, implicit volatilities, and the $\alpha$ exponent of anomalous diffusion, offering a novel approach to measuring the volatility of stock quotations. Additionally, this volatility measure allows us to classify assets into three types of stocks. Like the beta coefficient, companies are classified according to whether the beta is greater than, equal to, or less than one. Thus, knowing if it has a higher, equal, or lower risk than the market is important.

The main point is that the presented model for estimating the $\alpha$ exponent is performed based on synthetic data where noise is added. So, the provided predictions try to discount the presence of noise.

We have seen that for long time periods, these time series undergo a subdiffusion process, and the ATTM is particularly useful for modeling financial markets because it allows for the dynamic nature of volatility, especially over extended periods. This aligns with our results, where the probability of the ATTM model increases as the time window widens, reinforcing the idea that stock returns are better described by a process where diffusion rates vary randomly over time.

Besides, we have examined the impact of several specific events on the volatility exponent and

how the exponent behaves in less volatile sectors, such as Consumer Staples, following a subdiffusion process, while those from higher-risk sectors undergo a superdiffusion process, such as energy in the last few years.

Finally, we believe that future research, such as analyzing which of these three volatility measures better predicts future volatility or which volatility measure can better classify sectors, could shed more light on a better estimation of the market behavior and can also benefit from the recent progress of machine learning applied to physics.

## Code availability

The data sources and models used are available at [58]. As indicated, the models can be generated and trained following what is indicated in the Supplementary Material of [30]; see `https://github.com/AnDiChallenge/AnDi2020_TeamM_UPV-MAT`.

## Author contributions

Rubeń V. Arevalo, J. Alberto Conejero and Alfred Peris: Conceptualization, Methodology, Writing – original draft; Rubeń V. Arevalo: Data curation; Rubeń V. Arevalo and Ò. Garibo-i-Orts: Formal analysis; Rubeń V. Arevalo: Investigation, Validation. All authors have revised and approved the submission.

## Use of AI tools declaration

As described in Section 4, we have used a machine learning model to estimate the anomalous diffusion exponent. We have not used any other artificial intelligence (AI) tool to write this article.

## Acknowledgments

The third author was supported by MCIN/AEI/10.13039/501100011033/FEDER, UE, Projects PID2019-105011GBI00 and PID2022-139449NB-I00, and by Generalitat Valenciana, Project PROMETEU/2021/070. The second and third authors were supported by the European Union - NextGenerationEU, ANDHI project CPP2021-008994.

## Conflict of interest

The authors declare no conflict of interest in this paper.

## References

1. S. J. Brown, W. N. Goetzmann, A. Kumar, The Dow Theory: William Peter Hamilton's track record reconsidered, *J. Financ.*, **53** (1998), 1311–133. https://doi.org/10.1111/0022-1082.00054
2. O. B. Sezer, M. U. Gudelek, A. M. Ozbayoglu, Financial time series forecasting with deep learning: A systematic literature review: 2005–2019, *Appl. Soft Comput.*, **90** (2020), 106181. https://doi.org/10.1016/j.asoc.2020.106181

3. B. M. Henrique, V. A. Sobreiro, H. Kimura, Literature review: Machine learning techniques applied to financial market prediction, *Expert Sys. Appl.*, **124** (2019), 226–251. https://doi.org/10.1016/j.eswa.2019.01.012
4. G. Marti, F. Nielsen, M. Bińkowski, P. Donnat, A review of two decades of correlations, hierarchies, networks and clustering in financial markets, *Progress in information geometry: Theory and applications*, (2021), 245–274. https://doi.org/10.1007/978-3-030-65459-7_10
5. H. N. Bhandari, B. Rimal, N. R. Pokhrel, R. Rimal, K. R. Dahal, R. K. Khatri, Predicting stock market index using LSTM, *Machine Learn. Appl.*, **9** (2022), 100320. https://doi.org/10.1016/j.mlwa.2022.100320
6. E. F. Fama, Efficient capital markets: A review of theory and empirical work, *J. Financ.*, **25** (1970), 383–417. https://doi.org/10.2307/2325486
7. C. Meier, Adaptive Market Efficiency: Review of Recent Empirical Evidence on the Persistence of Stock Market Anomalies, *Rev. Integr. Bus. Econ. Res.*, **3** (2014), 268–280.
8. C. Chiarella, X.-Z. He, C. S. Nikitopoulos, Stochastic Processes for Asset Price Modelling, *Derivative Security Pricing: Techniques, Methods and Applications*, (2015), 7–36. https://doi.org/10.1007/978-3-662-45906-5_2
9. R. Brown, On the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies, *Edinburgh New Philos. J.*, **5** (1828), 358–371. https://doi.org/10.1080/14786442808674769
10. A. Einstein, On the motion of small particles suspended in liquids at rest required by the molecular-kinetic theory of heat, *Ann. Phys.*, **17** (1905), 208.
11. J. Perrin, Mouvement brownien et réalité moléculaire, *Ann. Chim. Phys.*, **18** (1909), 1–114.
12. R. C. Merton, Theory of rational option pricing, *Bell J. Econ. Manage. Sci.*, **4** (1973), 141–183. https://doi.org/10.2307/3003143
13. J. C. Hull, *Options, futures, and other derivatives*, Pearson Education, 2023.
14. Z.-C. Deng, J.-N. Yu, L. Yang, An inverse problem of determining the implied volatility in option pricing, *J. Math. Anal. Appl.*, **340** (2008), 16–31. https://doi.org/10.1016/j.jmaa.2007.07.075
15. B. Nabubie, S. Wang, Numerical techniques for determining implied volatility in option pricing, *J. Comp. Appl. Math.*, **422** (2023), 114913. https://doi.org/10.1016/j.cam.2022.114913
16. S. Li, A new formula for computing implied volatility, *Appl. Math. Comp.*, **170** (2005), 611–625. https://doi.org/10.1016/j.amc.2004.12.034
17. W. G. Hallerbach, An improved estimator for Black-Scholes-Merton implied volatility, *ERIM Report Series*, **54** (2004), 17. https://doi.org/10.2139/ssrn.567721
18. D. Stefanica, R. Radoicic, An explicit implied volatility formula, *Int. J. Theor. Appl. Finance*, **20** (2017), 1750048. https://doi.org/10.2139/ssrn.2908494
19. R. Yang, S. Du, J. Huang, Y. Zhang, A stock volatility prediction using hybrid machine learning models, *International Conference on Cyber Security, Artificial Intelligence, and Digital Economy (CSAIDE 2022)*, **12330** (2022), 449–454. https://doi.org/10.1117/12.2647212

20. P. Sadorsky, Machine learning methods for volatility forecasting in financial markets, *J. Risk Financial Manag.*, **15** (2022), 1–18.

21. K. S. Moon, H. Kim, Performance of deep learning in prediction of stock market volatility, *Econ. Comput. Econ. Cybern. Stud. Res.*, **53** (2019), 77–92. https://doi.org/10.24818/18423264/53.2.19.05

22. G. Di-Giorgi, R. Salas, R. Avaria, C. Ubal, H. Rosas, R. Torres, Volatility forecasting using deep recurrent neural networks as GARCH models, *Computational Statistics*, (2023), 1–27. https://doi.org/10.1007/s00180-023-01349-1

23. L. Alfonso, R. Mansilla, C. Terrero-Escalante, On the scaling of the distribution of daily price fluctuations in Mexican financial market index, *Phys. A*, **391** (2011), 2990–2996. https://doi.org/10.1016/j.physa.2012.01.023

24. F. De Domenico, G. Livan, G. Montagna, O. Nicrosini, Modeling and simulation of financial returns under non-Gaussian distributions, *Phys. A*, **622** (2023), 128886. https://doi.org/10.1016/j.physa.2023.128886

25. B. B. Mandelbrot, *The variation of certain speculative prices*, Springer New York, 1997. https://doi.org/10.1007/978-1-4757-2763-0_14

26. T. A. Waigh, N. Korabel, Heterogeneous anomalous transport in cellular and molecular biology, *Rep. Prog. Phys.*, **86** (2023), 126601. https://doi.org/10.1088/1361-6633/ad058f

27. L. Goiriz, R. Ruiz, Ò. Garibo-i-Orts, J. A. Conejero, G. Rodrigo, A variant-dependent molecular clock with anomalous diffusion models SARSCoV- 2 evolution in humans, *Proc. Natl. Acad. Sci. U.S.A.*, **120** (2023), e2303578120. https://doi.org/10.1073/pnas.2303578120

28. S. Manrubia, J. A. Cuesta, Physics of diffusion in viral genome evolution, *Proc. Natl. Acad. Sci. U.S.A.*, **120** (2023), e2310999120. https://doi.org/10.1073/pnas.2310999120

29. K. M. Graczyk, D. Strzelczyk, M. Matyka, Deep learning for diffusion in porous media, *Sci. Rep.*, **13** (2023), 9769. https://doi.org/10.1038/s41598-023-36466-w

30. G. Muñoz-Gil, G. Volpe, M. A. Garcia-March, E. Aghion, A. Argun, C. B. Hong, et al., Objective comparison of methods to decode anomalous diffusion, *Nat. Commun.*, **12** (2021), 6253. https://doi.org/10.1038/s41467-021-26320-w

31. G. Muñoz-Gil, H. Bachimanchi, J. Pineda, B. Midtvedt, M. Lewenstein, R. Metzler, et al., Quantitative evaluation of methods to analyze motion changes in single-particle experiments, *arXiv preprint arXiv:2311.18100*, (2023).

32. K. Ushida, A. Masuda, Chapter 11. General importance of anomalous diffusion in biological inhomogeneous systems, *Handai Nanophotonics*, **3** (2007), 175–188. https://doi.org/10.1016/S1574-0641(07)80016-7

33. J. Klafter, I. M. Sokolov, *First Steps in Random Walks: From Tools to Applications*, Oxford University Press, 2011. https://doi.org/10.1093/acprof:oso/9780199234868.001.0001

34. Y. Luo, C. Zeng, Negative friction and mobilities induced by friction fluctuation, *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **30** (2020). https://doi.org/10.1063/1.5144556

35. T. Huang, Y. Luo, C. Zeng, B.-Q. Ai, Anomalous transport tuned through stochastic resetting in the rugged energy landscape of a chaotic system with roughness, *Phys. Rev. E*, **106** (2022), 034208. https://doi.org/10.1103/PhysRevE.106.034208

36. Z. Ma, C. Zeng, W.-M. Liu, Relaxation time as early warning signal of avalanches in selforganizing systems, *Phys. Rev. Res.*, **6** (2024), 013013. https://doi.org/10.1103/PhysRevResearch.6.013013

37. A. Argun, A. Callegari, G. Volpe, *Simulation of Complex Systems*, 2021. https://doi.org/10.1088/978-0-7503-3843-1

38. J. E. Trinidad-Segovia, M. A. Sánchez-Granero, J. García-Pérez, Some comments on Hurst exponent and the long memory processes on capital markets, *Physica A*, **387** (2008), 5543–5551. https://doi.org/10.1016/j.physa.2008.05.053

39. M. Vogl, Hurst exponent dynamics of S&P 500 returns: Implications for market efficiency, long memory, multifractality, and financial crises predictability by application of a nonlinear dynamics analysis framework, *Chaos, Solitons & Fractals*, **166** (2023), 112884. https://doi.org/10.1016/j.chaos.2022.112884

40. R. Asif, M. Frömmel, Testing Long memory in exchange rates and its implications for the adaptive market hypothesis, *Physica A*, **593** (2022), 126871. https://doi.org/10.1016/j.physa.2022.126871

41. A. F. Perold, The capital asset pricing model, *J. Econ. Persp.*, **18** (2004), 3–24. https://doi.org/10.1257/0895330042162340

42. I. M. Sokolov, J. Klafter, A. Blumen, Fractional kinetics, *Phys. Today*, **55** (2002), 48–54. https://doi.org/10.1063/1.1535007

43. B. B. Mandelbrot, J. W. Van Ness, Fractional brownian motions, fractional noises and applications, *SIAM Rev.*, **10** (1968), 422–437. https://doi.org/10.1137/1010093

44. A. Argun, G. Volpe, S. Bo, Classification, inference and segmentation of anomalous diffusion with recurrent neural networks, *J. Phys. A*, **54** (2021), 294003. https://doi.org/10.1088/1751-8121/ac070a

45. Ò. Garibo-i-Orts, A. Baeza-Bosca, M. A. Garcia-March, J. A. Conejero, Efficient recurrent neural network methods for anomalously diffusing single particle short and noisy trajectories, *J. Phys. A*, **54** (2021), 504002. https://doi.org/10.1088/1751-8121/ac3707

46. D. Li, Q. Yao, Z. Huang, WaveNet-based deep neural networks for the characterization of anomalous diffusion (WADNet), *J. Phys. A*, **54** (2021), 404003. https://doi.org/10.1088/1751-8121/ac219c

47. C. Manzo, Extreme learning machine for the characterization of anomalous diffusion from single trajectories (AnDi-ELM), *J. Phys. A*, **54** (2021), 334002. https://doi.org/10.1088/1751-8121/ac13dd

48. E. A. Al-hada, X. Tang, W. Deng, Classification of stochastic processes by convolutional neural networks, *J. Phys. A*, **55** (2022), 274006. https://doi.org/10.1088/1751-8121/ac73c5

49. N. Firbas, Ò. Garibo-i-Orts, M. Á. Garcia-March, J. A. Conejero, Characterization of anomalous diffusion through convolutional transformers, *J. Phys. A*, **56** (2023), 014001. https://doi.org/10.1088/1751-8121/acafb3

50. Ò. Garibo-i-Orts, N. Firbas, L. Sebastiá, J. A. Conejero, Gramian angular fields for leveraging pretrained computer vision models with anomalous diffusion trajectories, *Phys. Rev. E*, **107** (2023), 034138. https://doi.org/10.1103/PhysRevE.107.034138

51. H. Verdier, M. Duval, F. Laurent, A. Cassé, C. L. Vestergaard, J. B. Masson, Learning physical properties of anomalous random walks using graph neural networks, *J. Phys. A*, **54** (2021), 234001. https://doi.org/10.1088/1751-8121/abfa45

52. M. A. Lozano, Ò. G. I. Orts, E. Piñol, M. Rebollo, K. Polotskaya, M. A. Garcia-March, et al., Open data science to fight COVID-19: winning the 500k XPRIZE Pandemic Response Challenge, *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, (2021), 384–399. https://doi.org/10.1007/978-3-030-86514-6_24

53. AnDiChallenge/ANDI datasets: Challenge 2020 release, 2021. https://doi.org/10.5281/zenodo.4775311

54. B. Y. Chang, P. Christoffersen, K. Jacobs, Market Skewness Risk and the Cross-Section of Stock Returns, *J. Financ. Econ.*, **107** (2009), 46–68. https://doi.org/10.2139/ssrn.1480332

55. P. Massignan, C. Manzo, J. A. Torreno-Pina, M. F. García-Parajo, M. Lewenstein, G. J. Lapeyre Jr, Nonergodic subdiffusion from Brownian motion in an inhomogeneous medium, *Phy. Rev. Lett.*, **112** (2014), 150603. https://doi.org/10.1103/PhysRevLett.112.150603

56. H. Scher, E. W. Montroll, Anomalous transit-time dispersion in amorphous solids, *Phys. Rev. B*, **12** (1975), 2455. https://doi.org/10.1103/PhysRevB.12.2455

57. J. Klafter, G. Zumofen, Lévy statistics in a Hamiltonian system, *Phys. Rev. E*, **49** (1994), 4873. https://doi.org/10.1103/PhysRevE.49.4873

58. Anomalous diffusion stocks article: release article, 2024. https://doi.org/10.5281/zenodo.14208749


AIMS Press