Universitat de les Illes Balears

FINAL DEGREE PROJECT

FINANCIAL MARKETS: A STATISTICAL ANALYSIS BASED ON THERMODYNAMICAL FORMALISM

Marçal Llopart i Molina

Degree in Physics

Faculty of Science

Academic Year 2021-22

FINANCIAL MARKETS: A STATISTICAL ANALYSIS BASED ON THERMODYNAMICAL FORMALISM

Marçal Llopart i Molina

Bachelor's Thesis

Faculty of Physics

University of the Balearic Islands

Academic Year 2021-22

Keywords:

Thermodynamics, Entropy, q-Statistics, TUR, Financial markets, market analysis, volatility, tail fitting.

Thesis Supervisor's Name: Dr. Pere Colet Rafecas

Co-Supervisor's Name: Dr. Rosa López Gonzalo

Abstract

English: In today's globalised world, understanding the dynamics of financial markets is key to preventing future crashes that could bring the system down to its knees. Traditionally, stock returns were modelled using gaussian distributions. However, real returns have large tails, and these models heavily underestimate them. We present a literature analysis and test several generalized distribution fittings to the logarithmic returns and its tails for three market indices. Another complication in financial modelling is dealing with volatility, which we try to connect to entropy as a measure of chaos or uncertainty. Its constraints from thermodynamical uncertainty relations (TUR) bounds both in literature and empirical data are studied and further applications are discussed.

Català: En el món globalitzat d'avui dia, la comprensió de la dinàmica dels mercats financers es clau per a prevenir futures crisis que podrien ferir greument el sistema. Tradicionalment, els rendiments de les accions es modelitzaven mitjançant distribucions gaussianes, però els rendiments reals tenen cues anormalment amples i aquests models els subestimen en gran mesura. Presentem una anàlisi de la literatura i testegem diversos ajustaments de distribució generalitzats per als rendiments logarítmics i les cues per a tres índexs de mercat. Una altra complicació en la modelització financera és el tractament de la volatilitat que intentem connectar amb l'entropia com a mesura de caos o incertesa. S'estudien les seves limitacions a partir dels límits de les relacions d'incertesa termodinàmiques (TUR) tant a la literatura com a les dades empíriques i es discuteixen altres aplicacions.

Castellano: En el mundo globalizado de hoy, la comprensión de la dinámica de los mercados financieros es clave para prevenir futuros choques que podrían dañar gravemente el sistema. Tradicionalmente, los rendimientos de las acciones se modelizaban mediante distribuciones gaussianas. Sin embargo, los rendimientos reales tienen colas anchas y estos modelos las subestiman notablemente. Presentamos un análisis de la literatura y testeamos varios ajustes de distribución generalizados para los rendimientos logarítmicos y sus colas para tres índices de mercado. Otra complicación en la modelización financiera es el tratamiento de la volatilidad, que intentamos conectar con la entropía como medida de caos o incertidumbre. Se estudian sus limitaciones a partir de los límites de las relaciones de incertidumbres termodinámicas (TUR) tanto en la literatura como en los datos empíricos y se discuten otras aplicaciones.

Index

1 Introduction

1.1 Complex systems

A fundamental problem in modern science is the study of so-called "complex" systems, in which many participants have a large degree of interaction which makes them difficult to model accurately because of the sheer computational cost. The applicability domain for these theories is incredibly broad, and ranges from the fields of Biology ([1],[2]), Chemistry([3], [4]) or Physics ([5],[6]) to sociology [7] or even finance, as we shall later see. These systems possess a number of interesting properties, since they are highly non-linear, correlated, and rife with emergent phenomena. However, the study of the system through each of its individual components and their interactions is not possible, and we must resort to a probabilistic approach, elucidating emergent macroscopic trends similarly to how we study thermodynamics from a Statistical Physics approach, employing the system's collective (average) variables as though it were a single entity evolving under a specific probability measure.

We can model the evolution of complex systems by what are known as *Stochastic Processes,* a random variable (or a collection thereof) of which the time evolution occurs over a probability space Ω through a certain *system sample path* $X_t(\omega)$ for each t value. One such process is the *Geometric Brownian Motion* (GBM), which follows the equation:

$$\frac{dY(t)}{Y(t)} = \sigma dW_t + \mu dt \ ; \quad (\sigma, \mu) \ const. ; \ S \in \mathcal{L}_2$$

Where dW_t is the differential for the Weiner process, the limit of random walks, such that the logarithm of the process will be a Brownian motion with drift, normally distributed with mean $x_0 + \mu t$ and variance $\sigma^2 t$, $\quad X(t) \sim \mathcal{N}(x_0 + \mu t, \sigma^2 t)$, where $X(0) = x_0$:

$$dX(t) = \sigma dW_t(t) + \mu dt$$

Here the σ, μ parameters give us the drift and volatility of the process.

1.2 Financial Markets

One of the modern backbones of the global economy are financial markets. Even though primitive forms of financial markets have existed since the Roman Republic (and then Empire), the powerhouse of ancient which took the torch from the Greeks and saw a precipitous evolution in the fields of banking and finance [8] which was picked up in Europe later, during the renaissance. However, these were far removed from the modern financial structures and exchanges we know today, of which the firstmost was established in Amsterdam c.1602 by the Dutch East India company [9].

In general, we refer to a financial market as any market where securities are traded, such as the various stock markets (New York Stock Exchange (NYSE), Nasdaq, London Stock Exchange (LSE), Euronext) to trade equity, bond markets for debt, Foreign Exchange(FX) to trade foreign currencies, and many others. *Securities* or *financial asset*s are defined as non-physical assets such as bonds, deposits or shares who are written on an underlying *real asset* (currency, property, etc). The main function of financial markets is that of providing a channel for the transfer of resources from agents with a surplus wishing to get returns on their assets towards agents in search of funds who are willing to pay a premium, doing this through the price discovery mechanism: bids are placed by potential buyers and matched with asking prices by sellers, and the price reflected in markets is that which in which the transactions go through, in other words, it is where supply and demand form an equilibrium. Today, this process happens in exchanges in a matter of nanoseconds [10]. While some markets are highly liquid (most securities markets fall into this category, such as the stock market or FX) others are more complicated to model due to large gaps and slow trading.

This price discovery mechanism is the result of the complex interactions between the millions of buyers and sellers which through the ensemble of their individual actions determine the evolution of the system, and so it could be useful to try treating it using the mathematical framework of statistical mechanics and complex systems and checking if different properties are sustained in a financial-market context.

Already in 1900, Louis Bachelier realized this in his PhD. Thesis [11] which even though badly received at the time, is now regarded as the first modern work on financial mathematics and explores the study of markets based on probability theory, postulating unpredictable (random) fluctuations around the equilibrium prices that are essentially what we know now as Brownian motion processes. In further works [12,13] he expanded these results and defined concepts such as Markov processes and path equations like Langevin's [14], therefore laying the foundation and inspiration for the works of Robert C. Merton, Fisher Black and Myron Scholes in option pricing, resulting in the Black-Scholes [15] and the Black-Scholes-Merton [16] models (to price European and American options, respectively) which form one of the eminent results of financial mathematics to date.

An option is a contract between two parties which grants the holder of the contract the right (but not the obligation, a key detail that distinguishes options from other future-trading instruments, such as futures contracts) to purchase or sell an underlying asset at a specified price and time in the future. In this regard, there are European style options, which can only be exercised on expiry, and American-style options, which can be exercised at any time prior. An options contract must always specify:

1. The underlying asset
2. The time at which the option holder can exercise his rights to buy or sell the underlying asset, which is usually denoted as T. It is the expiration date, or expiry.
3. The accorded price at which the underlying is sold/bought if the option is exercised, which is called the strike price, K.

The brilliance of these models lies in the simplicity and beauty of the resulting option pricing formulae. The revolution they brought about in the finance sector cannot be overstated: In 1973 the Chicago Board Options Exchange (CBOE) is founded, by 1977 1.1 million such contracts are traded and already in 1984 100 million are being annually traded [17].

I will expand on the BS model because it provides the perfect platform to understand the problem we aim to study. The key observation is that this model takes the underlying asset to follow a GBM, the equation of which we have seen above, so that its logarithmic return, $\log\,(Y_{t+1}/Y_t)$ (where Y_t is the price of the asset at time t) assumed to come exclusively from appreciation, is normally distributed. And from there we can convert it into a heat-diffusion equation[18], to solve in the general case. *Fig 0* shows several simulations for normal random walks and stock evolution simulations as GBM for varying parameters of drift and volatility. However, real returns do not show this gaussian behaviour.

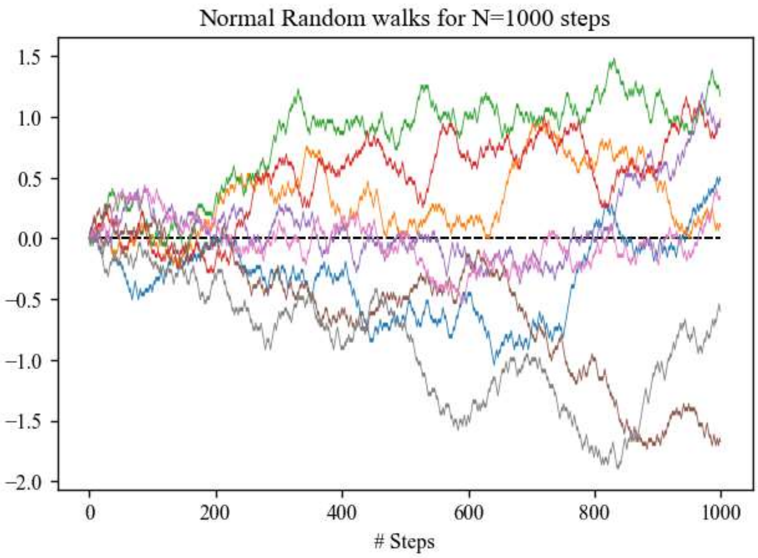
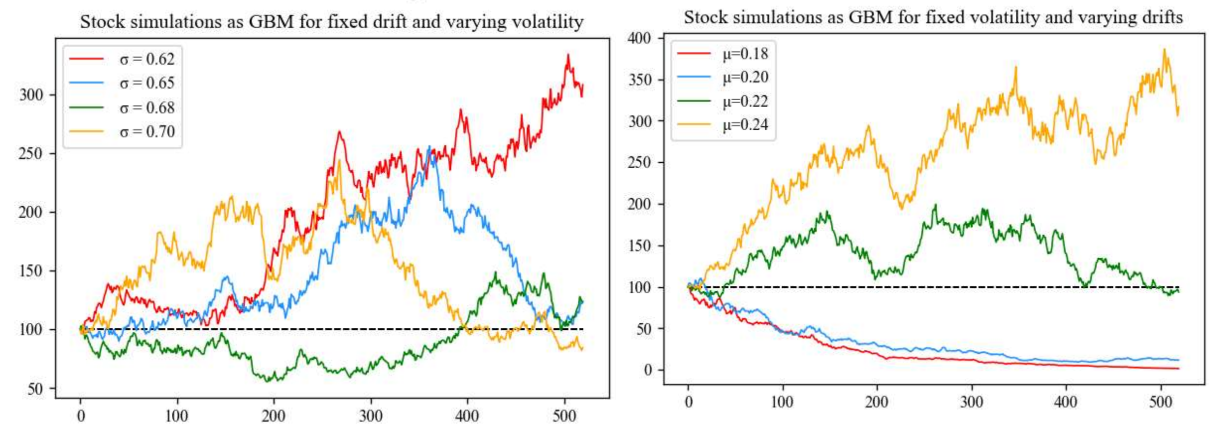


Figure 0 : Normal random walk simulations for N=1000 steps, and stock simulations as GBM (N=500 steps) for fixed drift (volatility) and varying volatility(drift) respectively (left, bottom left , bottom right, respectively)



In obtaining such results, as we have said (and most analytical ones in the field) two key approximations must be made: First, we typically assume that stock prices follow a GBM, or equivalently, that logarithmic stock returns are distributed according to a gaussian structure. Normal distributions do well to approximate the distributions of log-returns around the mean, but its tails heavily underestimate the likelihood of large-sigma events. In financial markets, behaviour tends to be non-stationary due to economic growth and the effects of inflation and the real behaviour is also non-equilibrium (meaning that the efficient market hypothesis does not hold). This means that, due to the tail risk, large crashes such as the ones that occurred in 2007 or 2020, which fall outside six normal distributions of the mean should happen only twice in a billion events if they followed these Gaussian distributions. We will explore the possibility of using several generalized gaussian distributions, the q-Gaussian family of distributions, to see if their characteristic kurtotic tails better approximate the real behaviour we observe in our empirical data sets.

The second approximation concerns the nature of the volatility σ of the underlying asset for the lifetime of the option, which in finance is typically associated with the variance or standard deviation of our data. The problem here is that financial time series are heterocedastic, namely, its volatility changes in time with periods of low volatility (in which markets are more or less stable) interrupted by periods of high volatility usually associated to crisis. When forecasting prices, we always need the *future* volatility, be it for the BS model or any other, but this is not an observable quantity: given the market now, we cannot know the exact value for σ at some future time t, only any given past value of it. This has no "easy" solution. It is important to know what the volatility of the asset is because it measures the uncertainty contained in the system: a high volatility is synonymous to a state of low informative diffusion, and vice versa, since the more information is contained, spread through agents, and thus reflected in the market, the more certain it becomes. To this effect, many workable solutions have been proposed, such as stochastic volatility in ARCH and GARCH models [19]. However, from empirical data we may gauge the volatility for any given time series, and thus we will try see if another measure of uncertainty can be used: Entropy. We want to observe the behaviour of entropy and its evolution in financial markets and compare it with the volatility since this could provide us with a useful alternative in situations where the complexity of a system could be estimated but not its fluctuation amplitudes.

1.3 Entropy

The concept of entropy originates in the field of Thermodynamics and is first defined by Clausius [20] to be interpreted as a measure of randomness or chaos of a system, or its degrees of freedom. It is central to the definition of the Second Law, which is one of the most important results to date in physics and has the most equivalent formulations. Its results have been extended to a variety of fields, such as information theory, probability theory, and finance. Several entropies are used in financial mathematics, such as the Shannon entropy or the Tsallis entropy we will present three such entropies to compare the results obtained by measuring each.

1.3.1 Shannon Entropy

First proposed by C. Shannon in 1948 [21], the Shannon entropy (sometimes also called Boltzmann-Gibbs (BG) or Boltzmann-Gibbs-Shannon (BGS) entropy) measures the information or uncertainty inherent in a random variable's possible paths, and for some probability measure p such that $X = \{p_i\}$ is defined as [22,23] for discrete-time:

$$H(X) = S_n(X) \equiv -\sum_{i=1}^{n} p_i \log p_i \tag{1}$$

Where X is the random variable, and $0 \ln 0 = 0$. Of course, the probabilities are taken to be normalised, so that $\sum_{i=1}^{n} p_i = 1$. The entropy for a continuous-time variable is simply:

$$H(X) = S_n(X) \equiv -\int_{-\infty}^{+\infty} p(x) \log p(x)\, dx\,, \qquad \int_{-\infty}^{+\infty} p(x) dx = 1 \tag{2}$$

These equations are valid for any non-negative, normalised probability distribution p.

1.3.2 Tsallis Entropy

A generalisation of Boltzmann-Gibbs (BG) statistics, called nonextensive statistics or q-statistics was proposed by C. Tsallis in 1988 [24] and expanded upon in subsequent papers for the next several decades [25-31].The starting point of this model is the *q-entropy,* which is really the same as the entropy proposed in information theory by Havrda & Charvát in 1967 [32]:

$$H_q(X) = S_q \equiv k \frac{1 - \sum_{i=1}^{W} p_i^q}{q-1} \quad (q \in \mathbb{R} \ \wedge k \geq 0) \tag{3}$$

And for a continuous random variable, we have:

$$S_q[p] = \frac{1}{q-1}\left(1 - \int_{-\infty}^{\infty} p^q(x) dx\right) \tag{4}$$

The B-G entropy is a subcase of this entropy which is recovered in the $q \to 1$ limit. We note that the real parameter q constitutes the degree of non-extensivity of the system, and in general this is not an extensive entropy, contrary to what we would expect. This is because in the standard formulation for statistical mechanics, interactions are assumed to be short-ranged and fast decaying with distance, so that extensivity arises as a property, but in reality, does not exist due to the long-range forces (such as gravity) that are present. When the system becomes non-ergodic, the central limit theorem that ensures the distribution is an exponential or gaussian (and therefore, ensuring the system is in equilibrium) does not hold, and the systems are rather in some quasi-stationarity represented by a q-exponential or q-gaussian distribution [33-36]. We will test this hypothesis later by fitting q-Gaussian distributions to the logarithmic returns using several q parameters to explore which could better approximate the real data.

1.3.3 Rényi Entropy

First proposed by Alfréd Rényi [37], the entropy that bears his name is also a generalisation of the Shannon entropy, and as such recovers the usual entropy at the $\alpha \to 1$ limit. As before, we can write this entropy for a discrete and a continuous random variable, such that we have:

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \left(\sum_{i=1}^{n} p_i^\alpha \right), \qquad \forall \alpha \geq 0 \cap \alpha \neq 1 \tag{5}$$

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \int_0^\infty p^\alpha(x) dx \ , \qquad \forall \alpha \geq 0 \cap \alpha \neq 1 \tag{6}$$

Where we label it H_α to distinguish it from the other generalised entropy, H_q. This entropy was derived by Rényi using Fadeev's postulates [38] for the Shannon entropy and Erdös's result [39] for additive number-theoretical functions such that H_α would be the most generalized entropy function that still preserves the additivity of entropy for independent events. This has been used in applications for the Principle of Maximum Entropy such as calibrating the prices for options [40]. We will also measure this entropy to compare its results against those that are obtained when employing the previous entropies (Shannon, Tsallis).

1.3.4 Kullback – Leiber divergence, relative entropy

The relative entropy or K-L divergence, introduced in the 50s by Kullback and Leiber[41,42], is a sort of statistical distance measure that allows us to gauge the difference between two probability distributions P,Q. Essentially, it tells us the information divergence between the two observed distributions, by measuring the average bits needed to encode events in p under model q, and so if both distributions have almost the same information content, the relative entropy is almost zero: The closer the informative content is, the lower the K-L divergence, such that:

$$D_{KL}(P \parallel Q) = \sum_x P(x) \log \left(\frac{P(x)}{Q(x)} \right) \tag{7}$$

For some two discrete probability distributions over a joint probability space. It therefore tells us how far two distributions are from each other, and we can use it to gauge the adequacy of fitted distributions to our data.

1.4 Thermodynamic Uncertainty Relations

One of the most recently developing areas of research is that of stochastic thermodynamics. In the words of Udo Seifert (2012) "Stochastic thermodynamics [...] systematically provides a framework for extending the notions of classical thermodynamics like work, heat, and entropy production to the level of individual trajectories of well-defined non-equilibrium ensembles" [43] from this field of research, which now supports a broad and growing literature, especially in its biological and biomolecular applications arise the Thermodynamic

Uncertainty Relationships (TUR). First presented by Barato and Seifert in 2015, they showed that for any steady-state Markov process the dispersion of the output generated by a biomolecular system is constrained by its entropic cost [44].Similar results are obtained in [45]. Several papers followed, expanding the scope of these relations, such as [46] to bound the efficiency of molecular motors, or presenting bounds outside the gaussian regime (so-called "universal", with large regions of validity) [47]. The bounds presented in the latter paper are of the parabolic and exponential kind and are further tightened subsequently for several kind of processes ranging from non-Markovian processes to stationary or periodic Markov processes [48-52]. Applications for this kind of formalism has been wide [53-61] and new connections and applicability domains are established on a constant basis. One of the aims of this work is to establish whether the assumption of such a bound holds when looking at empirical market data, from which some crude form of entropy evolution and production will be estimated.
In general, TUR relate the average and variance of some thermodynamic quantity to the entropy production as:

$$\langle f(\Delta \mathrm{S})\rangle \geq \frac{\langle \phi \rangle^2}{\mathrm{Var}[\phi]} \tag{8}$$

For some stochastic process $x \in \{x_i\}, i = 1, \ldots, N$ with some probability distribution $P(x)$ such that $\langle \cdot \rangle$ is the average with respect to P, $f(\Delta \mathrm{S})$ is some function for the entropy production $\Delta \mathrm{S}$ and ϕ is typically some time-asymmetric current [62-67]. One of the main ideas behind the TUR is that the relative precision for some measurement, given by the second term $\frac{\langle \phi \rangle^2}{\mathrm{Var}[\phi]}$ is bounded by the entropy production: It is not possible for our measurements to be precise beyond some upper point. Generally, we take $f(\Delta \mathrm{S})$ to be an increasing positive function like $f(\Delta S) = \frac{\langle \Delta \mathrm{S} \rangle}{2}$, such that higher relative precision has the requirement of a higher entropy production as well, which is energetically costly. In this view, the bound can be viewed as an energetic bound as well.
For a discrete-time Markov chain subject to time-symmetric external driving, we have that the TUR is the so-called "General" TUR [68-70]

$$\frac{\langle \Delta \mathrm{S} \rangle}{2} \geq \frac{\langle \phi \rangle^2}{\mathrm{Var}[\phi]} \rightarrow \frac{e^{\langle \Delta \mathrm{S} \rangle} - 1}{2} \geq \frac{\langle \phi \rangle^2}{\mathrm{Var}[\phi]} \tag{9}$$

The interest of this particular case is that it is a bound valid for any dynamic, even non-Markovian, for any time-antisymmetric observable.

TURs are, in fact, a specific realization for the Linear Fluctuation-Response Inequalities (LFRI) case for specific perturbation choices of the broader Fluctuation-Response Inequalities (FRIs) commonly associated with generalizations of the second law of Thermodynamics, which were first presented by Evans et al in 1993 [71] and which pose a solution to Loschmidt's paradox (time-reversal symmetry of most physical laws). In short, these inequalities show that even though for nonequilibrium systems the probability at any time of system's entropy flow not being what would be expected via the second law of thermodynamics is nonzero, it decreases exponentially with the size of the ensemble, so that in the macroscopic limit the expected second law is recovered, and the ensemble average is never negative.

However, while these developments are recent, the idea that thermodynamical quantities could form uncertainty relationships like the bounds obtained through the Heisenberg uncertainty principle are not in any way recent, being first suggested by Bohr and other scientists in the late 1900s [72-77]. An analysis of such bounds can be found in [43].

2 Methods

The following analysis has been conducted with the help of several python scripts we have programmed using the high-level programming language Python, as well as some common modules for different analysis functionalities [79-83]. We split the analysis into two main components, the static analysis and the evolutive analysis. We will study and compare an index Exchange Traded Fund (ETF) (SPY) two stock market indexes (Ibex-35, Nikkei-225).

The SPY (SPDR S&P 500 ETF) is traded on the NYSE Arca and is an exchange-traded index fund tracking the S&P 500 market index, which is composed of 500 of the largest companies traded in stock exchanges inside the USA, such as Apple (AAPL), Alphabet (GOOGL), Chevron(CVX) or Ford (F). The S&P 500 market index evolution is often taken as being representative of the financial well-being of the USA economy at large, due to the wide representation of major players in all main economic sectors. Therefore, it is strongly subjected to national economic policy in the USA and allows us to gain insight into its financial health.

The Ibex-35 is the stock market index for the 35 most liquid stocks traded in the *Bolsa de Madrid*. It contains some of the largest public Spanish companies such as Iberdrola (IBE), Inditex (ITX) or Santander (SAN). It offers an example of a southern-European occidental market, with typically strong correlations to global tendencies due to the strong dependence of the Spanish economy on tourism, leisure, and travel industries.

Finally, the Nikkei 225 market index tracks the evolution of 225 of the largest publicly traded Japanese companies, such as Mitsubishi or Yamaha, from the Tokyo stock exchange, which is the third largest exchange in the world by market capitalisation. This index presents interesting characteristics, deviating heavily from typical index dynamics, and usually showing strong shock responses to unpredictable natural disasters such as tsunamis or earthquakes.

2.1 Overall time series analysis

The analysis hereafter described has been performed on each of the three indicators mentioned above. Firstly, data has been imported from the yahoo finance database for the previous 10 years and daily granularity and handled using the pandas statistical module. Only closing prices have been employed as a measure of a stock's price on a given day, and from these the logarithmic daily returns, also called log-returns have been calculated as $\log\left(\frac{Y_{t+1}}{Y_t}\right)$ $\forall t$ for a total of 2518, 2448 & 2557 datapoints, respectively.

We further calculate some statistical parameters such as the standard deviation, mean and kurtosis of said logarithmic returns over the whole time series. The log returns are then distributed into an $N = 250$ bin histogram according to their relative frequency and their entropy is measured through the Shannon, Tsallis and Renyi measures which we have defined above.

Then, several curves are fitted to the logarithmic returns, using the statistical parameters we calculated, starting with a gaussian distribution, with a pdf that can be written as:

$$N(\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{11}$$

Finally, multiple q-Gaussians are fitted to the data. We want to test if these are better fits for the logarithmic returns, in particular at the tails, where we know our usual distributions cannot approximate them well. q-Gaussian distributions have a pdf given by:

$$QG(x) = Z\left(1 + \left(\frac{q-1}{3-q}\right)\left(\frac{x-\mu}{\sigma}\right)^2\right)^{\frac{1}{q-1}} \tag{12}$$

Where Z is defined as

$$Z \equiv \left(\frac{\mathrm{q}-1}{(3-\mathrm{q})\sigma^2}\right)^{-\frac{1}{2}} \cdot \Gamma\left(\frac{1}{2}\right) \cdot \Gamma\left(\frac{1}{\mathrm{q}-1} - \frac{1}{2}\right) \cdot \left(\Gamma\left(\frac{1}{\mathrm{q}-1}\right)\right)^{-1} \tag{13}$$

And we fit distributions for $q \in \{1.25, 1.5, 2, 2.5\}$ values.

For each distribution fitted, we now measure the entropy using the Shannon and Rényi ($\alpha = 0.75$) formulas. For Tsallis entropy, we calculate the entropy contained in every distribution fitted, as well as the entropy of the original time series for all values of $q \in [1,4]$. Finally, all of the previous results have been plotted using several data visualisation modules such as matplotlib and seaborn on python.

2.2 Statistical Analysis on time windows

For the second section of our analysis, we extend our data import to the past 20 years (2002-2022), and we separate it into 180-day data windows for a total of 45 data sets for each indicator. After computing the log returns, we distribute for each of these sets a $N = 25$ bin histogram and from those probabilities the Shannon and Rényi ($\alpha = 0.75$) entropies are calculated. We compare these to the volatility in each period identified as the variance of log returns to see if they move in synchrony. For each window the entropy production is calculated as the series $\Delta S = S_{t+1} - S_t$ $\forall t$. We then graph the entropy evolution as well as the scaled or fractional entropy production $\Delta S_{scaled} = \frac{S_{t+1}-S_t}{S_t}$ to better compare the different measures. The statistical moments are also calculated for the whole period, as well as the functions $\frac{\langle \Delta S \rangle}{2}, \frac{e^{\langle \Delta S \rangle}-1}{2}$ that appear in TUR taking $\langle \Delta S \rangle$ as the average of all entropy differences and the violation of such bounds is checked.

3 Results

3.1 Overall time series analysis

The evolution for the close prices of the three indicators is graphed in *Fig.*1. We see that this is a mostly growing period for both Nikkei 225 and SPY, with upward tendency and very few negative months or even weeks, with the latter having the steadiest growth and the former following a more erratic path. On the other hand, Ibex 35 rises until mid-2015 and then starts to decline slightly, crashing heavily in 2020.

The most notable price drop happens from the end of February 2020 until the first week of April in the same year. Amid the concerns in 2019 that an economic slowdown or even recession was incoming [78], the outbreak of the COVID-19 pandemic decimated the global economy and caused a 30% + fall for most major market index funds, with the SPY, Nikkei 225 and Ibex 35 being no exceptions. The strong dependency of the Spanish economy (which Ibex is a gauge of) to tourism, travel, and other service-related industries, the worse affected by the pandemic, make it so that the crash registered by this index is the strongest by far, and recovery the slowest due to the lingering restrictions on travel for the subsequent quarters. However, a quick recovery far above pre-COVID-19 levels is observed in both Nikkei and SPY, fuelled by the bullish attitude of individual investors and the quick expected recovery post-pandemic which peaks before a small decline in recent weeks.

We see in *Fig.2* larger return amplitudes for all three indicators in this time period, which reflects the higher volatility we would expect in such circumstances due to the wild price fluctuations. The largest positive % changes happen on the 13&24 March 2020, closing +9.29&+9.3% respectively for SPY and, and the largest drops, known as The Black Mondays (9,16 March 2020) and Black Thursday (12 March 2020) register 7.60%,11.98% and 9.51% drops respectively (SPY). The same movements are observed in the other two indicators, and these form the large-sigma events we observe in the asymmetric tails in *Fig.3*'s histogram.

We observe the period between 2016 and 2018 has the most stable and smallest returns for the SPY, but the opposite is true for the Nikkei index, which presents remarkably high fluctuations in this period, in which the Ibex has a single large drop and stable evolutions later. In general, the SPY presents the most stable returns, while the Nikkei 225 has consistently high fluctuations across the whole time series, and Ibex 35 presents large single events and consistent moderate fluctuations. For all three, some fluctuation clustering is observed around 2012,2016,2020 & 2022.

We have represented the N=250 bin histogram edges in *Fig.3*, which allows us to observe the kurtotic nature of returns as well as the asymmetry in the far end of tails, which is negatively skewed (as in, the negative tails extend further than the positive ones), contrary to smaller events which are more likely to be positive, something particularly notable for the Ibex 35 index.

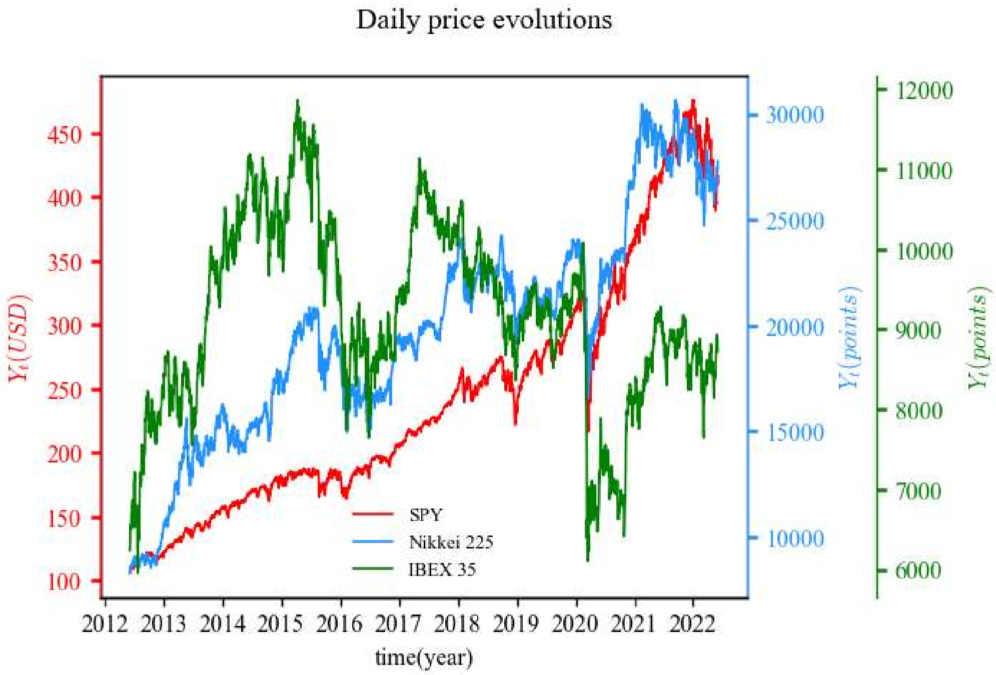


Figure 1: Evolution of the stock during the past 10 years, taking the closing price for each trading day for a total of 2518, 2448 & 2557 data points. All three evolutions are plotted in a single graph for ease of comparison, with each scale set using the same colour as the curve it concerns.

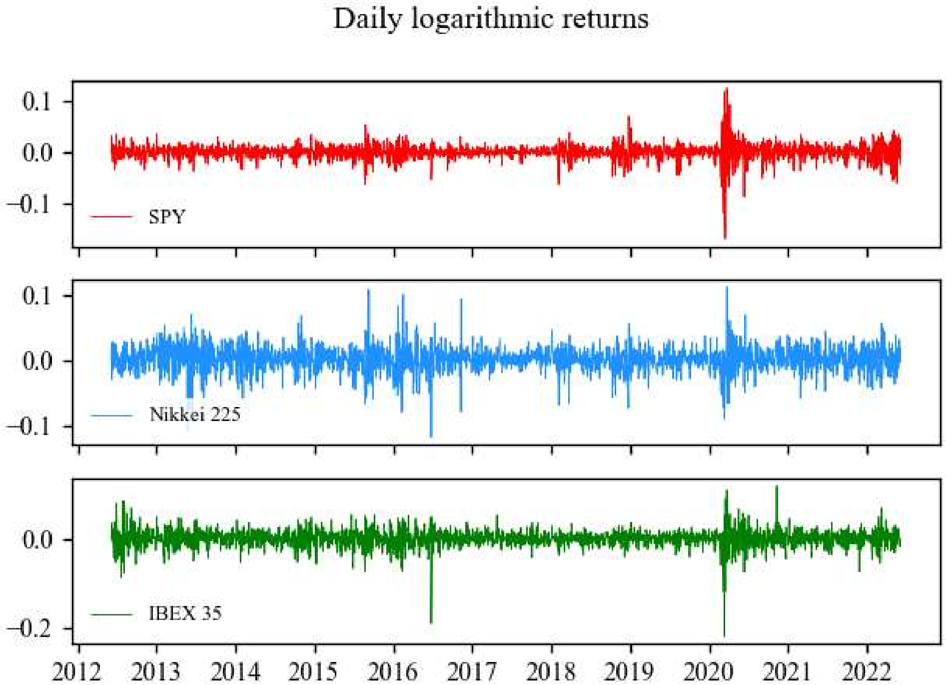


Figure 2: Evolution each of the indicator's logarithmic returns during the past 10 years, taking the logarithms of the closing price for each trading day for a total of 2518, 2448 & 2557 data points.

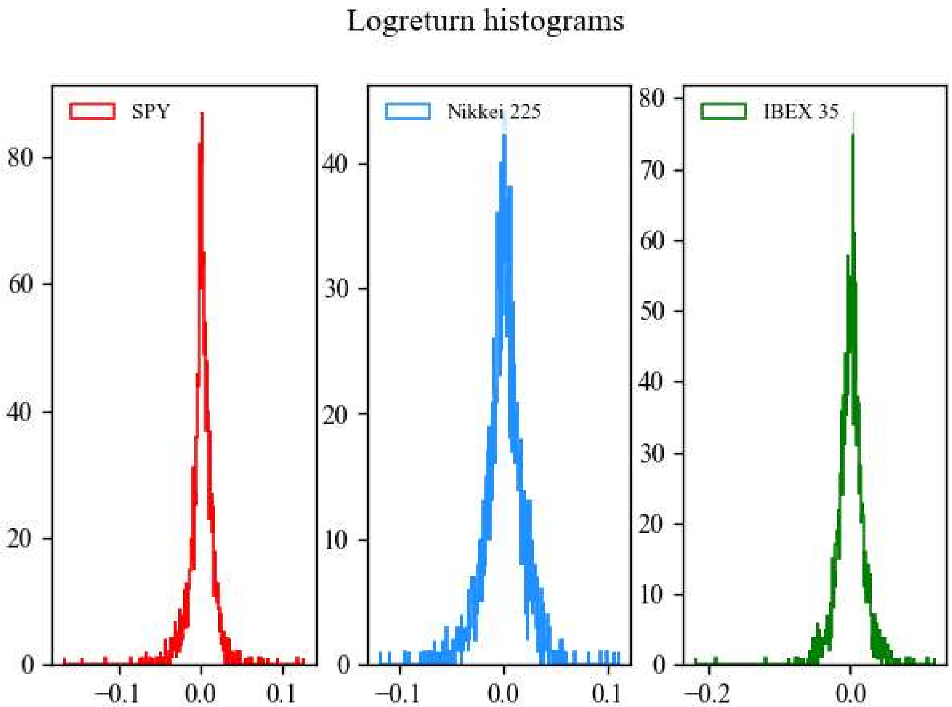


Figure 3: Distribution of the stock's logarithmic returns during the past 10 years, in which the kurtotic nature of the distribution is observed. The returns are distributed into an N=250 bin histogram and the edges are plotted.

We see in *Fig 4,5,6* the curves we have fitted to our logarithmic returns. Normal distributions underestimate the importance of the central event density and large sigma events, while at the same time overestimating the frequency of medium returns. On the other hand, we see that for q-distributions, the tails get progressively larger as a function of growing q values, which allows us to better model the returns. We see that Nikkei 225 returns have a broader distribution, with wide midsection and not such large tails, while both Ibex and SPY returns are far slenderer in the midsection and narrower width and fatter tails, which is much more "textbook" log-return behaviour. In *Fig 7,8* & 9, we can see a visual comparison for all the fits against the actual data, which allows us to better see how the tails compare between distributions. The "eye-tests" tells us the q-Gaussians with parameters $q = 2, 2.5$ are the better fitting ones, and so does the K-L divergence statistical distance measure we have calculated, which can be found in *Table 1*. For the Nikkei 225 Index, the best fit is that of the q = 2 Gaussian distribution, while for both SPY and Ibex 35 the best fits are those of a larger parameter value 2.5 Gaussian distribution.

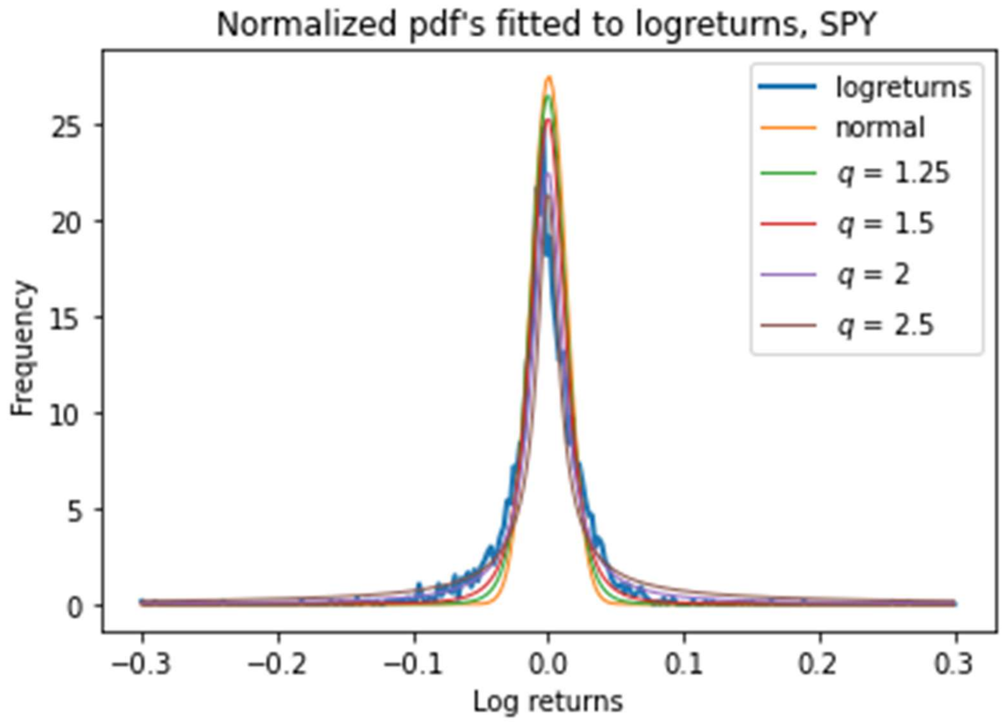


Figure 4: Log return distribution vs fitted distributions for SPY returns (10 years)

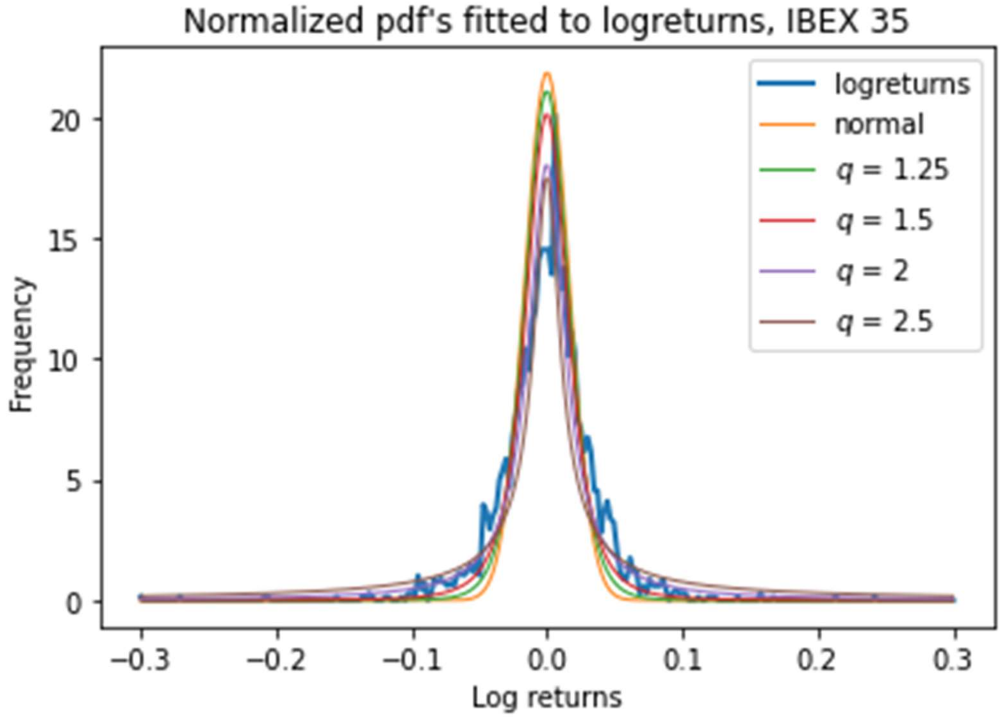


Figure 5: Log return distribution vs fitted distributions for Ibex 35 returns (10 years)

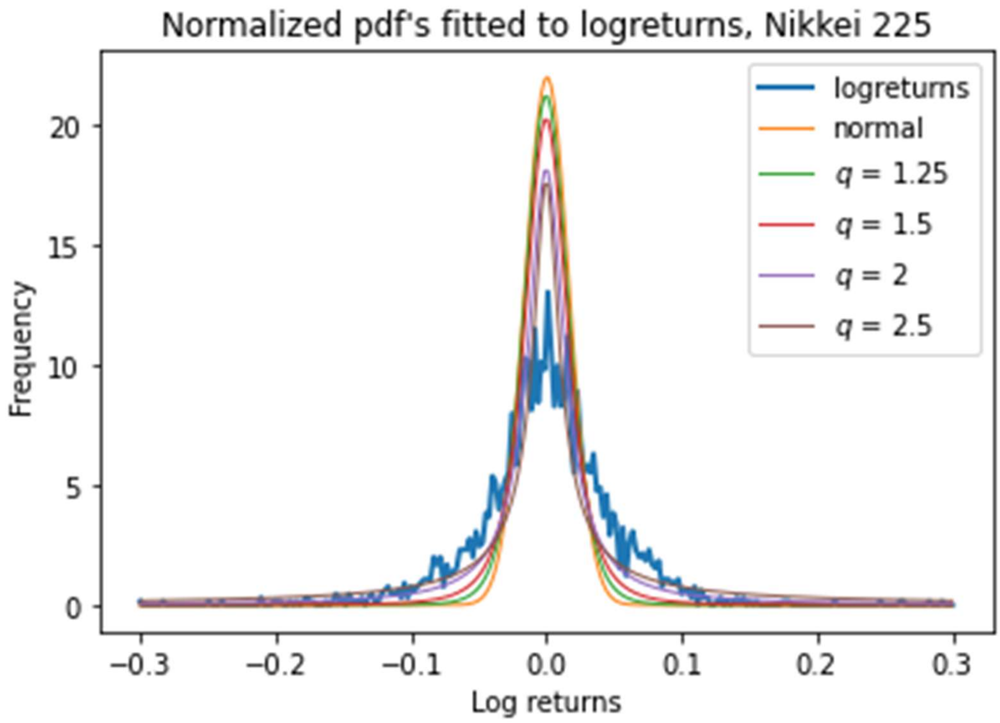


Figure 6: Log return distribution vs fitted distributions for Nikkei 225 returns (10 years)

	Gaussian	$q = 1.25$ G	$q = 1.5$ G	$q = 2$ G	$q = 2.5$ G
SPY	2773.99	1149.61	806.54	577.51	528.96
Ibex 35	5913.24	2032.32	1358.42	906.09	764.85
Nikkei 225	1263.82	314.43	158.86	80.959	90.190

Table 1: KL divergence statistic distance or relative entropy used to test the goodness of the fits we have employed. I have shortened q-Gaussian to q-G.

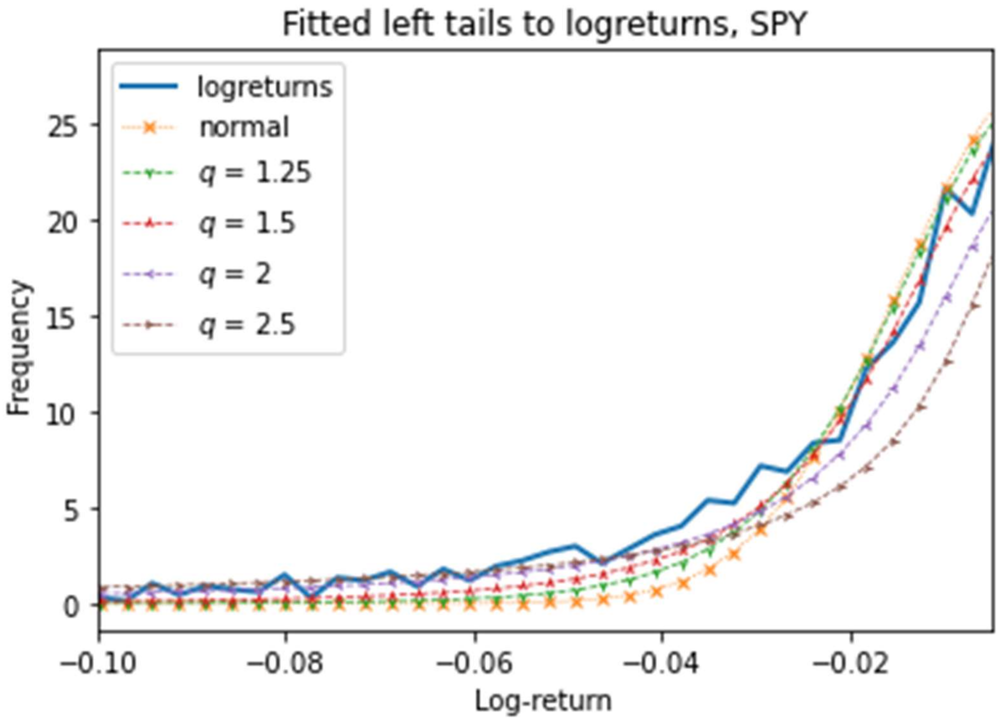


Figure 7: Comparison between all distributions and the return data (blue), for the 10-year SPY returns, zoomed on the left tails of the distributions. We see that the central region is well approximated by the Gaussian and q = 1.25 distributions, whereas the further tail-ends are much better approximated by q=2, q=2.5 Gaussian distributions.

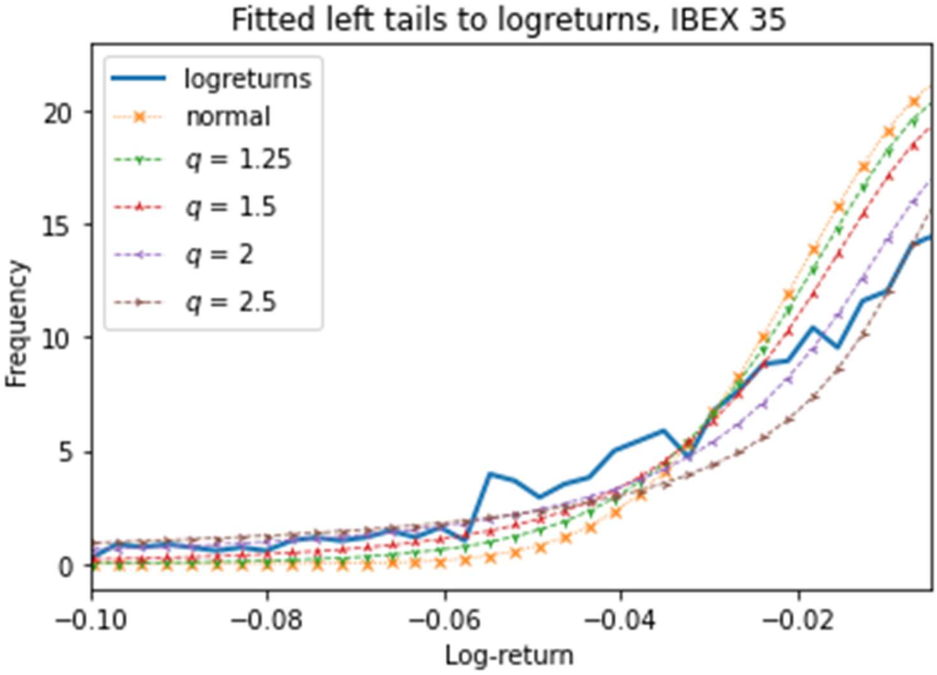


Figure 8: Comparison between all distributions and the return data (blue), for the 10-year Ibex 35 returns, zoomed on the left tails of the distributions. We see that the central region is not too well approximated by all except q = 2.5 distributions, and the further tail-ends are much better approximated, in particular by q=2, and q=2.5 Gaussian distributions again.

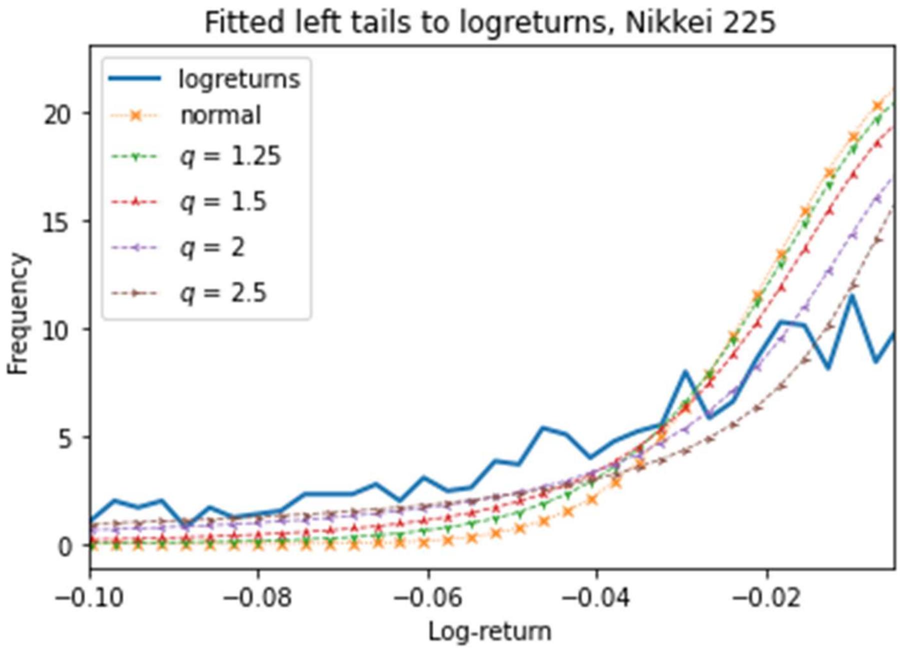


Figure 9: Comparison between all distributions and the return data (blue), for the 10-year Nikkei 225 returns, zoomed on the left tails of the distributions. We see that the central region is very badly approximated by all distributions, and the further tail-ends are better approximated but still heavily underestimate the far events.

3.2 Statistical analysis on time windows

The second part of our analysis consists of a study around the evolution of entropy contained in our three time series. As has explained in the methods section, we run 180-day windows through the data to analyse the evolution of entropy.
The results are presented in *Figs.10, 11* & *12*, where we can see that IBEX seems to have the highest entropic variations over the whole timeframe, which can be partly because we have run our analysis over 20+ years, and therefore the events surrounding 2007 are included, which destabilised the Spanish economy and could cause strong variation in the dispersion of the logarithmic returns (which is what entropy is an indicator of). Later swings could indicate confirmation of the strong dependence of this economy on commodities that affect the travel and service, as well as the construction sector. The SPY is the most consistent, indicating a stable, quasi-stationary economy excepting some large sigma events around the subprime and Covid crashes. Finally, Nikkei falls in between the two. It does not share the same pattern for evolution, having a small response to both subprime and covid crises but large swings in 2011 (when the Fukushima nuclear disaster occurred) and 2016.

It is notable to observe that the largest entropic drops are one-time events across all-time series; for instance, in the case of SPY, these would correspond, from left to right, to the 180-day windows starting on 06/04/2007, 12/09/2011, 07/02/2018 and 26/07/2020 respectively (*Fig.10).* The general evolutive patterns are common to all three entropic measures across the three analysed indicators, with the Rényi and Tsallis entropy showing smoother patterns with smaller jumps or changes.

However, correcting for the modulus of entropy, we can see in *Figs. 13, 14* & *15* that the fractional entropic change or entropic production is almost identical independently of the measure used to calculate it. In this regard, it does not seem like any of the measures pose particular advantages when conducting these kinds of analysis except as a means of verifying the coherence of results.

While we would expect a strong correlation between volatility and entropy, the periods with highest volatility do not correspond to those with markedly higher or lower entropies (Annex II). Low entropic contents indicate a small dispersion in the logarithmic returns and thus strong market correlations, where all behaviours tend in the same direction for the data in our time series since such a behaviour indicates the information needed to build the series is low.
However, we observe that strong entropic and volatility movements have some form of lagged or asynchronous correlations: In almost all cases across the three indicators, entropic drops precede or succeed volatility peaks by a semester, which may indicate that returns agglutinate before a period of strong volatility and vice versa at the start of a bear market

or bull market, in which sentiment is shared by investors and trends simplify after strong rises or sharp crashes (see 2007 movements), suggesting that whenever large shocks are observed in either the entropic or volatility curves, traders should hedge against near-future shocks in the other. Despite this, it is interesting to note that the intensity of events bears no correlation: Large swings in one do not translate into large swings for the other, which is to be expected since the entropy calculations only take into account the dispersion of the values or frequency thereof rather than the values themselves. *Table 2* Shows the Mean, standard deviation ,MVAR and $f(\langle \Delta \mathrm{S} \rangle)$ for the entire series from 2000-present.

	IBEX 35	Nikkei 225	SPY
Mean(μ)	-7.628 $\cdot 10^{-5}$	2.127 $\cdot 10^{-2}$	3.535 $\cdot 10^{-4}$
Standard deviation (σ)	2.100 $\cdot 10^{-2}$	1.101$\cdot 10^{-4}$	1.781 $\cdot 10^{-2}$
μ^2/σ^2	1.3234 $\cdot 10^{-5}$	2.679 $\cdot 10^{-5}$	4.079 $\cdot 10^{-4}$
$\langle \Delta \mathrm{S} \rangle/2$	-0.004038	-0.0028217	-0.0011693
$(e^{\langle \Delta \mathrm{S} \rangle} - 1)/2$	-0.004021	-0.0028137	-0.0011679

Table 2: Statistical parameters for the analysis conducted on the time series of log returns for the whole time series from 2000-present, where μ^2 indicates the squared mean of log returns over the period and σ^2 is the variance of said returns over the whole period as well. The $\langle \Delta \mathrm{S} \rangle$ function results shown here stem from the averages for the Shannon entropy. The numerical results for Tsallis and Renyi entropies are not shown here because they share the same behaviour as Shannon. The bottom two rows are the left hand sides of TUR shown in (9) and the third row is the right hand side, for each of the indicators.

We also find the bounds established by TUR are not violated for any data sets under our calculations, where we have computed $\langle \Delta \mathrm{S} \rangle$ for each indicator as the average of all entropic variations across the entire window series, and checked both bounds found in (9), but we must note that this average which is usually taken over the set of all realizations for the studied process is in this case calculated as a time-average over the entropy differences for the whole time series which are not necessarily equivalent. If we recall the underlying assumptions for such bounds establish processes which are asymmetric currents under time reversal, and so one usually has some kind of stochastic master equation or state equation from which energy, work , enthalpy and all manner of state functions can be derived. However, when analysing a financial process of this kind, all we have is empirical data and time reversibility becomes ill-defined: What is the time-reversed process to the evolution of ticker pricings? A price involution does not exist. There are other systems in which TUR have been well studied and hold as expected, such as some biological processes. The key difference is that in our system we cannot control a large number of realizations for the process with parameter variations or time reversal to study it: We only have a single system path realised in a singular time direction.

Further work could be conducted regarding this crucial aspect: Some measures of time-dependence could be inferred for the returns, from which a primitive form of time-direction could be conceived to test its reversal. In Annex III we see a matricial structure for joint probabilities of the log-returns at time t and lagged time t-1 that could be perhaps refined to infer the entropic production of this process. We see that it is not symmetrical, which indicates some underlying structure for the logreturns, meaning that the process should have nonzero entropy production, which is in accord with our results. It is also noted that the nonzero elements are slightly wider in diagonal directions, indicating that similar movements aggregate. The main problem with these kinds of approaches is that vast amounts of data is required: Sub-1min data is almost a necessity to obtain statistically valid results, and this kind of frequency is not in general openly available.

We must also note that the data availability for our purposes is already subpar: for each trading day we have only a single point, such that for any given 180-day window we get only ~ 123 events. *Fig 16* shows the dependence of entropy on binning frequency for SPY daily returns (5037 events, 20 years of trading), we see that Tsallis entropy seems to have the highest stability in front of binning changes; in general, below a 1:10 bin to data ratio the sensibility to binning frequency slows down but may be a factor to be considered when considering numerical results.

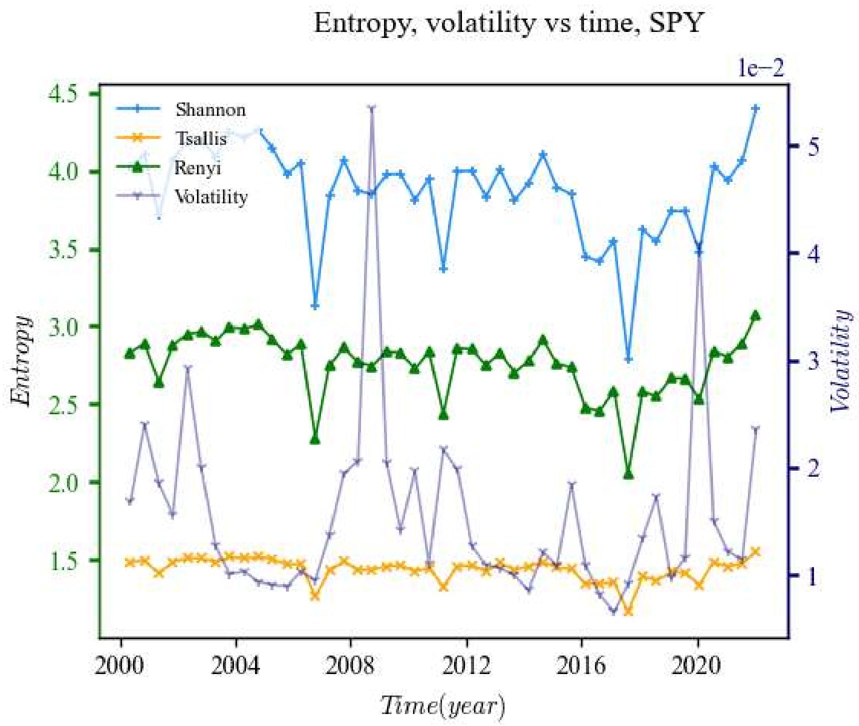


Figure 10: Evolution of the entropy values as measured under Shannon, Tsallis (q=1.5), Rényi ($\alpha = 0.75$) equations for SPY. In violet we can contrast the timeline for entropic movements with that of the variance in each window.

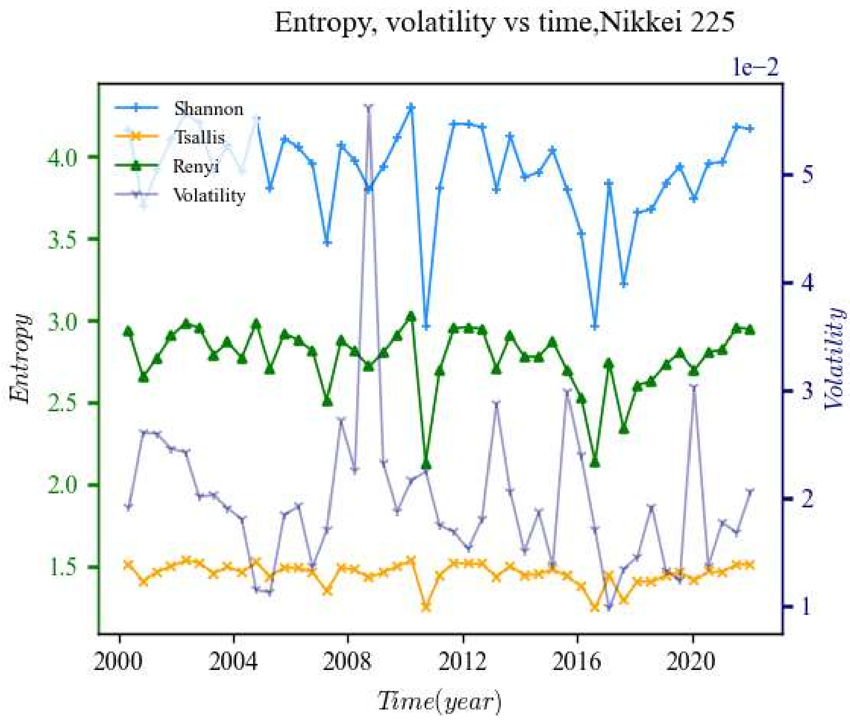


Figure 11: Evolution of the entropy values as measured under Shannon, Tsallis (q=1.5), Rényi ($\alpha = 0.75$) equations, for Nikkei-225 returns (45 events) In violet we can contrast the timeline for entropic movements with that of the variance in each window.

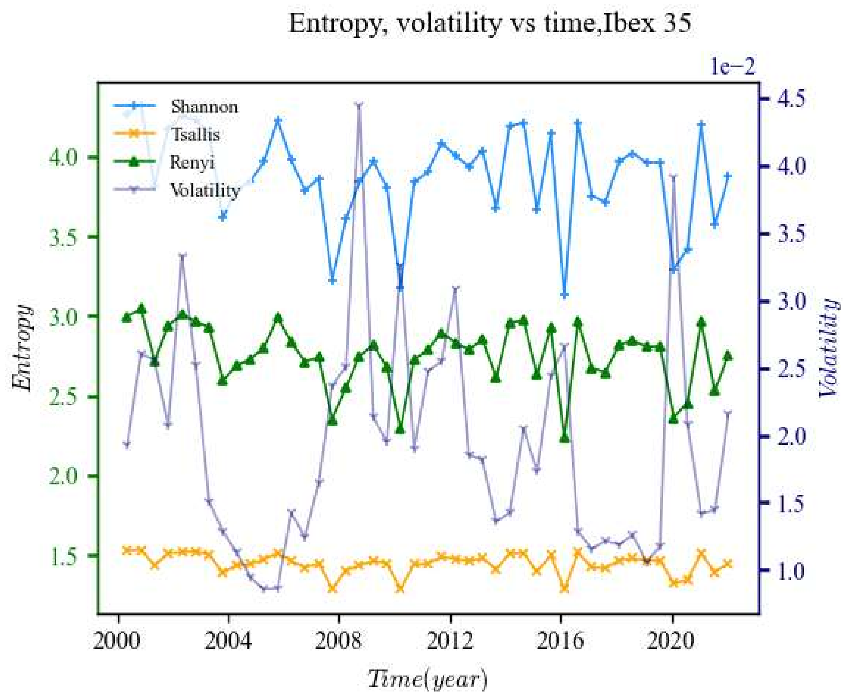


Figure 12: Evolution of the entropy values as measured under Shannon, Tsallis (q=1.5), Rényi ($\alpha = 0.75$) equations, for Ibex-35 returns (45 events) In violet we can contrast the timeline for entropic movements with that of the variance in each window.

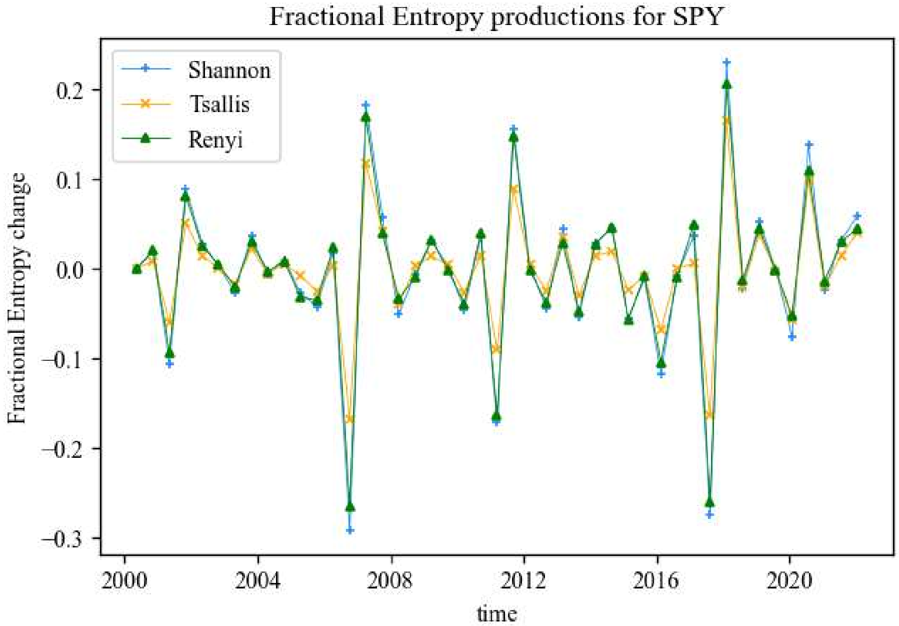


Figure 13: Fractional entropy productions for SPY inside each 180-day window change for 2000-present. It can be observed that Rényi entropy has some divergence from Tsallis & Shannon entropies, but these present almost identical changes with strong fluctuations in 2007, 2011,2018 & 2020

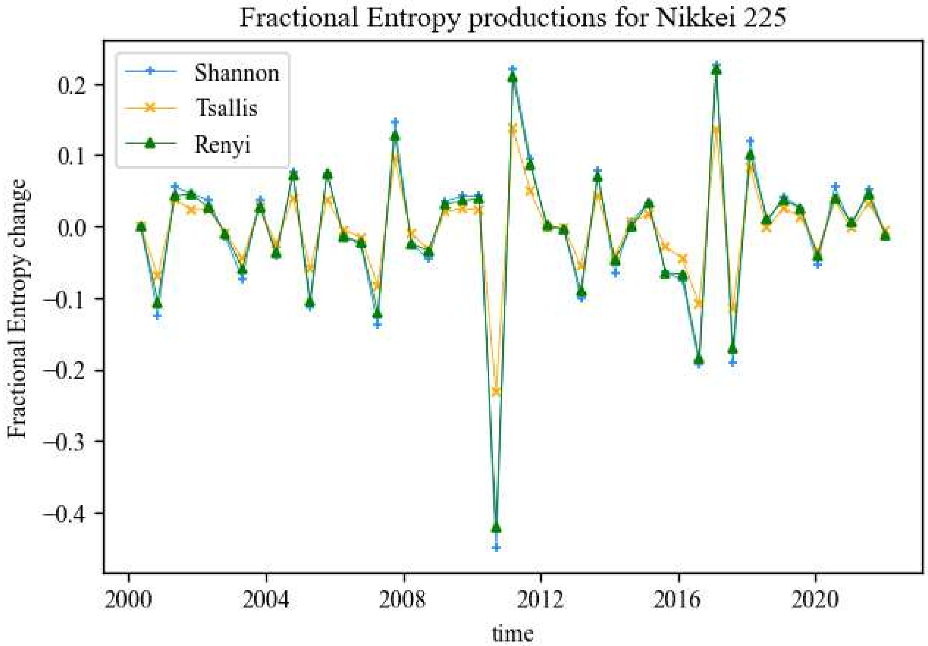


Figure 14: Fractional entropy productions for Nikkei 225 inside each 180-day window change for 2000-present. Contrary to the other indicators, only small fluctuations are observed in 2007 & 2020, but large ones are seen in 2011 & 2017.

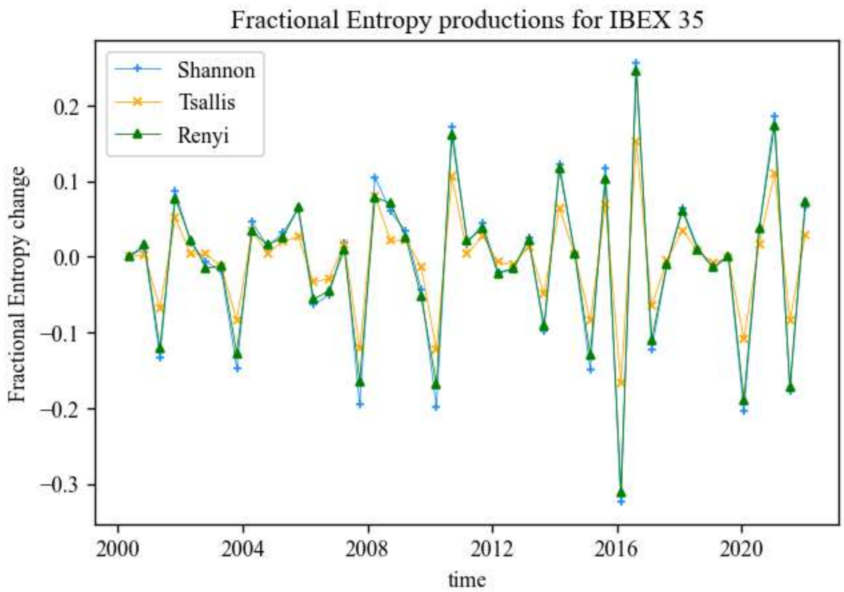


Figure 15: Fractional entropy productions for Ibex 35 inside each 180-day window change for 2000-present. Markedly large entropic swings characterize the whole time series, with no characteristic pattern observable. This is the highest fluctuating of the three.

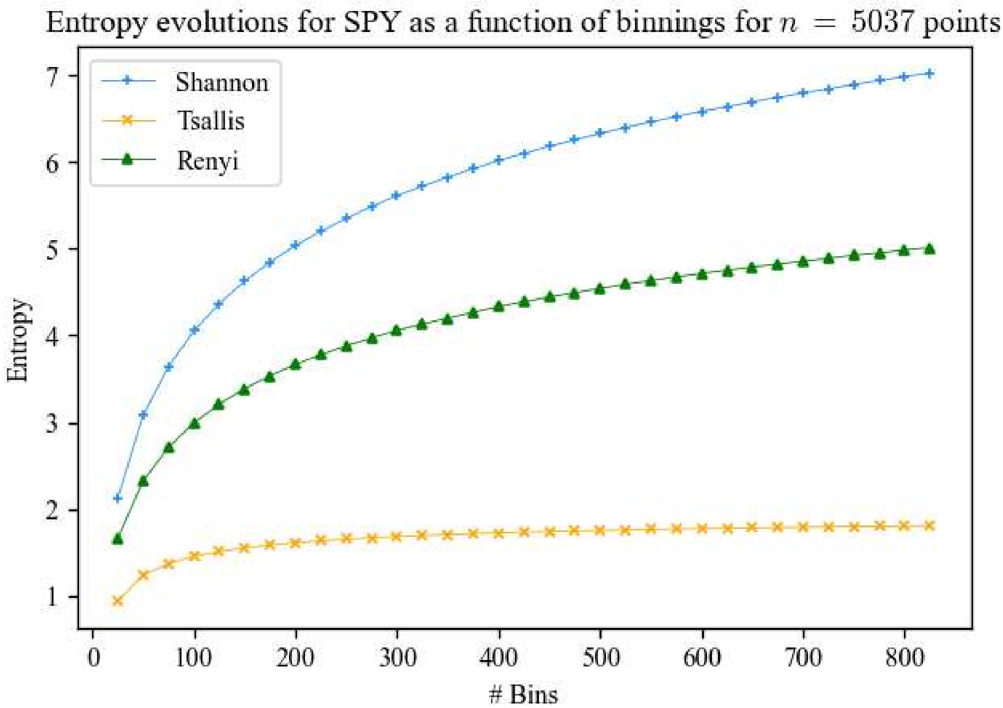


Figure 16: Dependence of entropy on the binning of data, for the past 20 years of logarithmic returns.

4 Conclusions

In this thesis, we have explored applications of concepts developed in the context of thermodynamics or statistical physics to the study of financial markets, taking empirical data from three market indicators (Ibex 35 - Spain, Nikkei 225 – Japan, SPY – Tracking S&P 500, USA) to test our hypothesis. An extensive literature review of the field has been conducted, and this thesis has also been used to develop the code needed for all the analysis, which totals just shy of 6000 lines of code written over the last year.

We first conducted an analysis for a static time series over the past 10 years, which we used to fit different generalized q-Gaussian distributions to the logarithmic returns to try and find better modelling options than normal distributions. In all our results, the $q = 2{,}2.5$ Gaussian distributions presents the better fits, and we see that the tails are much better approximated in this way. However, discrepancies still occur, especially with the Nikkei 225 returns, far less centralized and very erratic, such that even these q-distributions which possess larger tails underestimate them.

Then, we conducted an analysis from the year 2000-present taking semestral time series and analysing their entropy under three different measures: Shannon entropy, Tsallis entropy with parameter $q = 1.5$ and Rényi entropy with parameter $\alpha = 0.75$. We also measure the volatility and compare its movements to the evolution of entropy and find that almost everywhere we have lagged correlations: When substantial changes are seen in entropy, the volatility spikes up at a close, later time, and the other way around. This leads us to think that indeed entropy may be a good indicator crashes or other unstable market conditions. Further work could be done to measure whether this behaviour is observed only for large movements or holds also for any variation of entropy, which could be useful in predicting the future fluctuations of financial markets as a function of their complexity and vice versa.

Finally, we calculate a primitive form of entropy "production" or entropy difference between analysis windows. We check the bounds proposed in (9) and see that they do hold for all analysed cases, but further analysis should be conducted to determine if the results are conclusive, since the assumptions made in the formulations in such bounds are at their core energetic and can only be imprecisely translated into financial quantities. We calculate the joint probability matrix and see that it is not symmetrical, possibly corroborating the nonzero entropy productions we have found. Further work could be done by developing non-binning methods, since this process is what scrambles information and destroys any time label to calculate the probabilities.

5 References

[1] Barabási A-L and Oltvai Z N 2004 *Network biology: understanding the cell's functional organization*

[2] Dehmer M, Emmert-Streib F, Graber A and Salvador A 2011 *Network Biology: Methods in Systems Biology*

[3] Ashkenasy, Gonen, et al. "*Systems chemistry.*" Chemical Society Reviews

[4] Whitesides, George M., and Rustem F. Ismagilov. "*Complexity in chemistry.*"(1999)

[5] Sornette D 2011 *Critical phenomena in natural Sciences* (New York: Springer-Verlag)

[6] Mézard M, Parisi G and Virasoro M A 1987 *Spin Glass Theory and Beyond*

[7] Galam S 2012 Sociophysics: *A Physicists Modeling of Psycho - Political Phenomena*

[8] P. Temin, "*Financial Intermediation in the Early Roman Empire*", The Journal of Economic History, Vol. 64, No. 3,pp. 705-733 (2004)ç

[9] Braudel, Fernand (1983). *Civilization and capitalism 15th–18th century: The wheels of commerce.* New York : Harper & Row. ISBN 978-0060150914

[10] Terrence Hendershott; Pamela C. Moulton (2011). *Automation, speed, and stock market quality: The NYSE's Hybrid.*

[11] Bachelier, L. (1900a), *« Théorie de la spéculation »* (PDF), *Annales Scientifiques de l'École Normale Supérieure*, vol. 3, no. 17, pp. 21–86

[12] Bachelier, L. (1906). « *Théorie des probabilités continues* », Journal de Mathématiques Pures et Appliquées,2, 259-327.

[13] Bachelier, L. (1912). *« Calcul des probabilités »,* Gauthier-Villars, Paris.

[14] Langevin, P. (1908). *« Sur la théorie du mouvement brownien ».* Comptes Rendus de l'Académie des Sciences, 146, 530-533

[15] Black, Fisher S.; Scholes, Myron S.; *The Pricing of Options and Corporate Liabilities* Journal of Political Economy,

[16] Merton, Robert C.; *Theory of Rational Option Pricing*, The Bell Journal of Economics and Management Science, Vol. 4, No. 1 (Spring, 1973), pp. 141-183

[17] https://www.cboe.com/about/history

[18]https://www.math.tamu.edu/~stecher/425/Sp12/blackScholesHeatEquation.pdf

[19] *Lecture notes: Financial time series, ARCH and GARCH models* Piotr Fryzlewicz Department of Mathematics University of Bristol

[20] Clausius, R. *The Mechanical Theory of Heat*; McMillan and Co.: London, UK, 186

[21]Shannon, Claude E. (1948). *"A Mathematical Theory of Communication". Bell System Technical Journal*

[22] T.M. Cover, J.A. Thomas, *Elements of Information Theory* (John Wiley & Sons, New York, 2006)

[23] M. Hayashi, *Quantum Information Theory – Mathematical Foundation* (Springer, Berlin, 2017)

[24] C. Tsallis, *Possible generalization of Boltzmann-Gibbs statistics*, J. Stat. Phys. 52 (1988)

[25] E.M.F. Curado and C. Tsallis, *Generalized statistical mechanics: connection with thermodynamics*, J. Phys. A 24, L69-72 (1991).

[26] C. Tsallis, *Non-extensive thermostatistics: brief review and comments*, (1995).

[27] C. Tsallis, *Entropy*, Encyclopedia 2, 264-300 (MDPI Publishers, 2022)

[28] C. Tsallis, *Nonextensive statistics: Theoretical, experimental, and computational evidences and connections* (1999)

[29] C. Tsallis, *Nonextensive thermostatistics and fractals, Fractals* 3, 541 (1995).

[30] C. Tsallis, *Remarks on the nonuniversality of Boltzmann-Gibbs statistical mechanics, Fractals*
[31] C. Tsallis, *Comment on "Critique of q-entropy for thermal statistics"* by M. Nauenberg
[32] Havrda, J.; Charvát, F. (1967). "Quantification method of classification processes. Concept of structural α-entropy" (PDF). *Kybernetika*. 3 (1): 30–35.
[33] Tsallis C, Anteneodo C, Borland L and Osorio R 2003 Nonextensive statistical mechanics and economics Physica A 324 89
[34] Osorio R, Borland L and Tsallis C 2004 Distributions of high-frequency stock-market observables Nonextensive Entropy: Interdisciplinary Applications ed C Tsallis and M GellMann (New York: Oxford University Press)
[35] Tsallis C 2000 Introduction to Nonextensive Statistical Mechanics (New York: Springer)
[36] S. Drozdz, M. Forczeka, J. Kwapien a , P. Oswieçimkaa , R. Rak (2007) Stock market return distributions: From past to present
[37] Rényi, Alfréd (1961). "On measures of information and entropy" (*Proceedings of the fourth Berkeley Symposium on Mathematics, Statistics and Probability 1960*. pp. 547–561.
[38] D. K. FADEEV, "*Zum Begriff der Entropie einer endlichen Wahrscheinlichkeitsschemás*" Arbeiten zur Informationstheorie I, Berlin, Deutscher Verlag der Wissenschaften, 1957, pp. 85-90.
[39] P. ERDÖS, "On the distribution function of additive functions," Ann. of Math.,2nd Series. Vol. 47 (1946), pp. 1-20.
[40] Option price calibration from renyi entropy D.C. Brody, I.R.C. Buckley, I.C.Constantinou Article in Physics Letters A · July 2007 DOI: 10.1016/j.physleta.2007.01.088
[41] Kullback, S.; Leíble (1951). "On information and sufficiency". *Annals of Mathematical Statistics*. 22 (1): 79–86. doi:10.1214/aoms/1177729694. JSTOR 2236703. MR 0039968.
[42] *Kullback, S. (1959), Information Theory and Statistics, John Wiley & Sons.*
[43] U. Seifert, *Stochastic thermodynamics, fluctuation theorems, and molecular machines*, Rep. Prog. Phys. 75 (2012) 126001.
[44] Barato A C and Seifert U Phys. Rev. Lett. 114 158101 (2015)
[45] T. R. Gingrich, J. M. Horowitz, N. Perunov, and J. L. England, "*Dissipation bounds all steady-state current fluctuations*", Phys. Rev. Lett. 116, 120601 (2016).
[46] Pietzonka P, Barato A C and Seifert U (2016) J. Stat. Mech. 124004
[47] P. Pietzonka, A. C. Barato and U. Seifert Phys. Rev. E 93 052145 (2016)
[48] Polettini M, Lazarescu A and Esposito M 2016 Phys. Rev. E 94 052104
[49] Gingrich T R, Horowitz J M, Perunov N and England J L 2016 Phys. Rev. Lett. 116 120601
[50] C. Nardini and H. Touchette Eur. Phys. J. B 91 16 (2018)
[51] C. Maes 2017 Phys. Rev. Lett. 119 160601
[52] Gingrich T. R., Rotskoff G. M., and Horowitz J. M. J. Phys. A: Math. Gen 50 184004 (2017)
[53] Di Terlizzi, I. & Baiesi, M. Kinetic uncertainty relation. J. Phys. A 52, 02LT03 (2018).
[54] Garrahan, J. P. Simple bounds on fluctuations and uncertainty relations for first-passage times of counting observables. Phys. Rev. E 95, 032134 (2017).
[55] Gingrich, T. R. & Horowitz, J. M. Fundamental bounds on first passage time fluctuations for currents. Phys. Rev. Lett. 119, 170601 (2017).
[56] Guioth, J. & Lacoste, D. Thermodynamic bounds on equilibrium fluctuations of a global or local order parameter. Europhys. Lett. 115, 60007 (2016).
[57] Barato, A. C., Chetrite, R., Faggionato, A. & Gabrielli, D. Bounds on current fluctuations in periodically driven systems. New J. Phys. 20, 103023 (2018).

[58] Barato, A. C., Chetrite, R., Faggionato, A. & Gabrielli, D. A unifying picture of generalized thermodynamic uncertainty relations. J. Stat. Mech. Teor. Exp. 2019, 084017 (2019).
[59] Koyuk, T., Seifert, U. & Pietzonka, P. A generalization of the thermodynamic uncertainty relation to periodically driven systems. J. Phys. A 52, 02LT02 (2018).
[60] Pietzonka, P., Barato, A. C. & Seifert, U. Affinity-and topology-dependent bound on current fluctuations. J. Phys. A 49, 34LT01 (2016).
[61] Koyuk, T. & Seifert, U. Operationally accessible bounds on fluctuations and entropy production in periodically driven systems. Phys. Rev. Lett. 122, 230601 (2019)
[62] U. Seifert, "Stochastic thermodynamics: From principles to the cost of precision,"
[63] A. C. Barato and U. Seifert, "Thermodynamic uncertainty relation for biomolecular processes,"
[64] T. R. Gingrich, J. M. Horowitz, N. Perunov, and J. L. England, "Dissipation bounds all steady-state current fluctuations," Phys. Rev. Lett. 116, 120601 (2016).
[65] P. Pietzonka, F. Ritort, and U. Seifert, "Finite-time generalization of the thermodynamic uncertainty relation," Phys. Rev. E 96, 012101 (2017). (2018).
[66] J. M. Horowitz and T. R. Gingrich, "Proof of the finitetime thermodynamic uncertainty relation for steadystate currents," Phys. Rev. E 96, 020103(R) (2017)
[67] P. Pietzonka and U. Seifert, "Universal trade-off between power, efficiency, and constancy in steady-state heat engines," Phys. Rev. Lett. 120, 190602
[68] Hasegawa, Y. and Van Vu, T. arXiv:1902.06376 (2019)
[69] P. P. Potts. and P. Samuelsson arXiv:1904.04913 (2019).
[70] K. Proesmans and C. Van den Broeck, "Discrete-time thermodynamic uncertainty relation,"
[71] Evans, Denis J., Cohen, E. G. D.; Morriss, G. P. (1993). "*Probability of second law violations in shearing steady states*". *Physical Review Letters*. 71
[72] N. Bohr, Collected Works, J. Kalckar, ed. (North-Holland, Amsterdam, 1985), Vol. 6.
[73] B. Lavenda, *Statistical Physics: A Probabilistic Approach* (1991).
[74] B. Mandelbrot, Phys. Today 42, 71 (1989)
[75] J. Lindhard, "*The Lesson of Quantum Theory*", J. de Boer, E. Dal, and O. Ulfbeck, eds. (North-Holland, Amsterdam, 1986).
[76] B. Lavenda, Int. J. Theor. Phys. 26, 1069 (1987)
[77] B. Lavenda, Statistical Physics: A Probabilistic Approach (Wiley, New York, 1991).
[78] https://www.imf.org/external/pubs/ft/ar/2019/eng/
[79] https://matplotlib.org/
[80] https://seaborn.pydata.org/
[81] https://numpy.org/doc/stable/index.html
[82] https://scipy.org/
[83]https://pandas.pydata.org/

ANNEX 1: Entropies for SPY

In *Table 1* we have the entropies calculated through several equations (Shannon, Tsallis, Rényi) for every distribution fitted and the logarithmic returns for . This allows us to compare the information content of these distributions, and we see that the results are consistent across the three measures, with the $q = 2.5$ Gaussian having the largest entropy and the $q = 1$ Gaussian (regular Gaussian, for short) having only slightly the lowest one. Tsallis entropy presents the least spread results (amplitude about 11% from the smallest to the largest in relation to the average) with Shannon entropy presenting the more spread ones (36%). In *Fig.1* we have plotted the Tsallis entropy curves for all q values $q \in \{1,4\}$ for each of the distributions and data; we see the choice of generality parameter q does indeed affect our numerical results, but tendencies remain unchanged with the same distributions having the largest and lowest entropies.

	Shannon	Tsallis ($q = 1.5$)	Renyi
Log return data	5.429903610011825	1.6637745411868106	3.9125348418766404
Gaussian	4.482560172032928	1.5565287683482614	3.18243808983849
$q = 1.25$ Gaussian	4.695268769884641	1.579823072632985	3.37193002099043
$q = 1.5$ Gaussian	4.990994801583901	1.6088958986816582	3.6446100023598844
$q = 2$ Gaussian	5.7650312134482995	1.6782324204226204	4.2752348638182625
$q = 2.5$ Gaussian	6.382138167970462	1.7310937170434122	4.690511326769886

Table 1: Entropy calculated for all the fitted distributions, as well as the raw data for logarithmic returns.

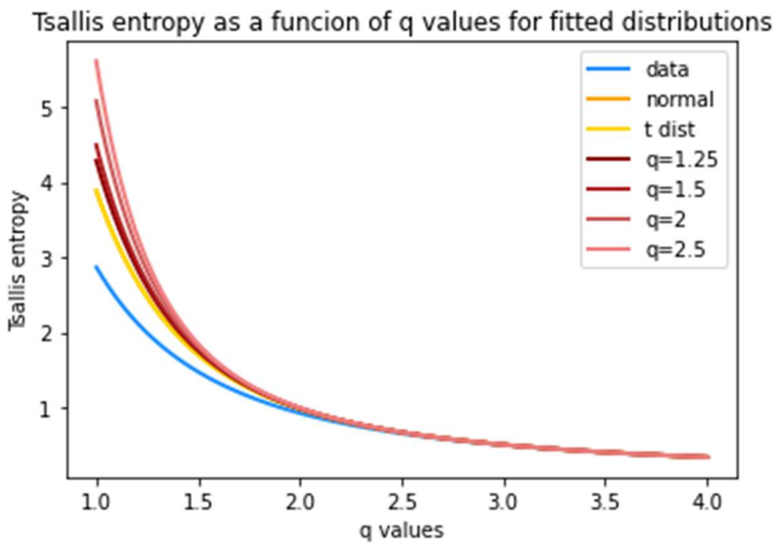


Fig 1: Tsallis Entropy curves for all distributions fitted

Annex II: Regressions for SPY

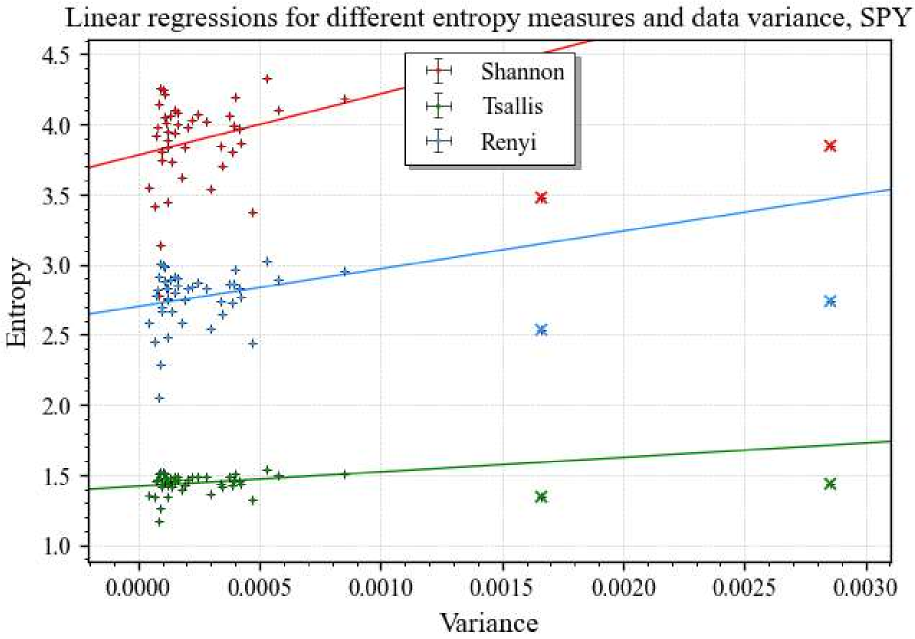


Figure 1: Linear regressions for the entropy vs variance for the three entropy measures.Outliers that have been eliminated (corresponding to the strongly non-linear periods of market crash in 2020 and 2008) are shown to be crossed out in the same colour as the data points in their respective sets.

	Shannon	Tsallis	Rényi
R coefficient	0.241978846173	0.244991518637	0.232238441649
R^2 coefficient	0.058553761995	0.060020844204	0.05393469378
A coefficient	3.7811353608022	1.4194873457451	2.701622173959
B coefficient	435.89962363533	102.77812729856	268.7146595496

Table 1 : Regression coefficients for the analysis represented in *Figure 10* . The linear regression equation is $y = A + Bx$.

ANNEX III : Log return Matrix

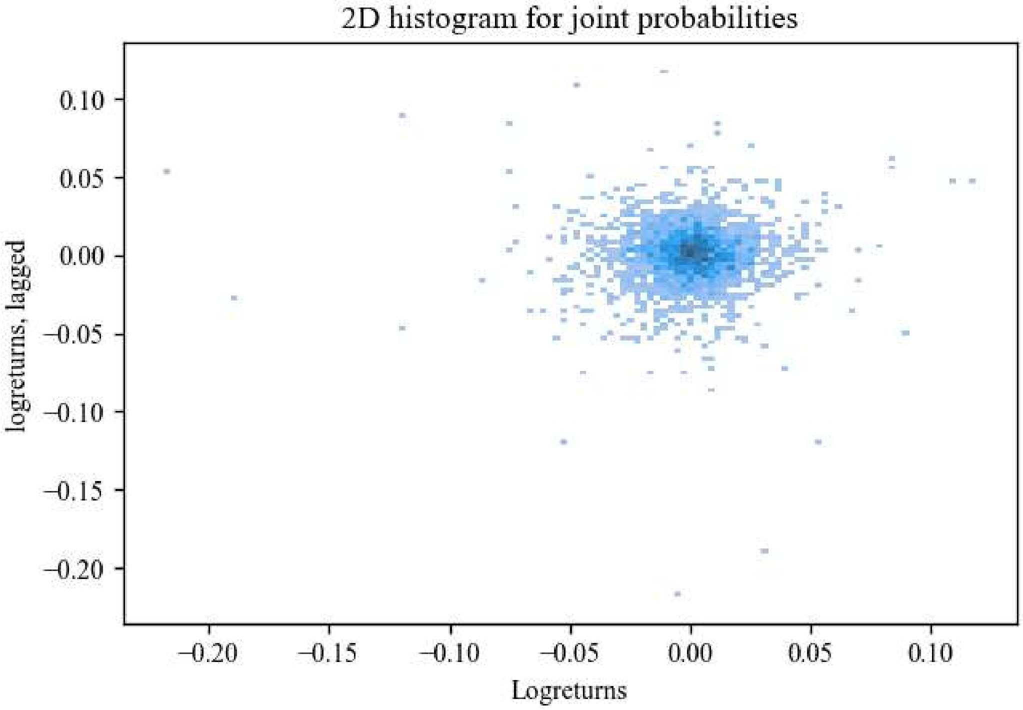


Fig 1:Logarithmic returns matrix. The colour gradient marks frequency of a certain value following some other in our time-series.