Bank i Kredyt 55(6), 2024, 731-758

BANK I KREDYT
BANK & CREDIT

Exploring stock markets dynamics: a two-dimensional entropy approach in return/volume space

Tomasz Kopczewski*, Łukasz Bil

Submitted: 9 April 2024. Accepted: 8 August 2024.

Abstract

This paper presents an entropy-based analysis of returns and trading volumes in stock markets. We introduce a measure of entropy in the return/volume space, leveraging Shannon's entropy, Theil's index, Relative Entropy, Tsallis distribution, and the Kullback-Leibler Divergence. We assess one- and two-dimensional returns and volume distributions, separately and jointly. This exploratory study aims to discover and understand patterns and relationships in data that are not yet well-defined in the literature. By exploring entropy measures, we identify mutual relations between returns and volume in financial data during global shocks such as the COVID-19 pandemic and the war in Ukraine. Revealing entropy changes in the return/volume space consistent with changes in the real economy allows for the inclusion of a new variable in machine learning algorithms that reflects the system's unpredictability.

Keywords: financial markets, return and trading volume, entropy, Kullback-Leibler Divergence, Tsallis distribution

JEL: G12, C58, C49

* Uniwersytet Warszawski, Wydział Nauk Ekonomicznych; e-mail: tkopczewski@wne.uw.edu.pl.

1. Introduction

The study began with a thought experiment: if we received financial time series samples without knowing the periods they represent, could we identify those from the pandemic versus the pre--pandemic period based on any unique statistic or characteristic? We aimed to see if the disruptions caused by the COVID-19 pandemic and war in Ukraine were reflected in financial market statistics and if they could be easily attributed to each period. It proved difficult to find a conclusive statistic that could reliably assign a sample to one of the two periods.

Lexical descriptions of financial markets during the pandemic – such as bizarre, weird, messy, unpredictable, and inconsistent with the real economy – suggest that entropy, as a measure of disorder, could be a reasonable criterion to distinguish these periods. Entropy has been used to study financial market uncertainty (Bentes, Menezes 2012), efficiency (Patra, Hiremath 2022), and behavioural efficiency (Dinga et al. 2021). According to the efficient markets hypothesis (Fama 1970), financial asset prices reflect all available information, but two basic variables characterize financial assets: price and trading volume. Thus, the entropy measure should consider both components. Observed price changes are crucial, but markets generate two simultaneous signals for investors: returns and volume. These two signals describe the market fully when considered together, as their mutual relations are inseparable (Bray 1981).

Most studies on the relationship between returns and trading volume focus on whether additional trading volume information improves price change forecasts. Most explanations are based on behavioural factors described in Simon's (1971) attention economic theory. Changes in trading volume attract investors' attention, adjusting their previous valuations based on social learning (Hou, Xiong, Peng 2009). Empirical evidence shows trading activity's influential role in future prices (Gervais, Kaniel, Mingelgrin 2001). A high-volume return premium exists in stock prices, directly related to economic fundamentals (Wang 2021), explained mainly by Merton's (1987) investor recognition hypothesis. Positive shocks to equity trading activity increase asset visibility, causing subsequent demand changes. It is worth noting that the authors of these works treat the relationship between volume and price change asymmetrically – increasing the volume has a positive impact on price changes. However, an increase in trading volume is also a basic indicator of the occurrence and bursting of speculative bubbles (Liao, Peng, Zhu 2022). There has never been a theory combining these two asymmetric approaches.

This article treats the returns and volume relationship differently from previous research. Assuming we treat returns and trading volume as one complex information system, we want to determine how unpredictable this system is or how much information is needed to determine the possible relationships between returns and volume. Entropy, as a measure of disorder, and relative entropy, as a measure of uncertainty in the two variables together (Zhou, Cai, Tong 2013), can answer this. Using the analogy of mixing two liquids (Aaronson, Carroll, Ouellette 2014), if the components are not mixed, system entropy is zero and fully predictable. In the financial system, zero entropy means returns and volume are in a constant, deterministic relationship. The more mixed the fluids, the more complicated the mutual relationships, and at maximum entropy, the relationship is purely random. In the financial market, this means infinite computational resources cannot improve the forecast based on these two components. We predict entropy and its components will be much higher during shock periods, but decrease in calmer times, with different mutual relations between the components of joint entropy in different periods. This decrease in disorder may seem to contradict the second law of thermodynamics, which states entropy should increase. However, financial markets are not isolated systems. During turbulent times, external information flows into financial markets, impacting prices and trading volumes and leading to more significant disorder. Keynes observed this unpredictability during crises influenced by external interactions and speculative behaviour (Neto 2022).

This study also has practical implications. In some financial markets, almost 80–90% of transactions are algorithmic (Hilbert, Darmon 2020). Changes in returns and volume provide signals for bots to make trading decisions. In most machine learning (ML) or artificial intelligence (AI)-based trading systems, bots operate as black boxes, where even their creators cannot interpret the complex models' relationships (Addy et al. 2024). The bot's decision-making process is purely computational, with only part of the calculations transformed into useful output. Using an analogy to Carnot efficiency (Clausius 1867), it's like a steam engine using only part of its thermal energy for work. Higher entropy means lower computation efficiency. We can extend this inference: the market can be seen as an ordering mechanism, similar to Maxwell's Demon, reducing information entropy, but requiring more external information/energy (Jones et al. 2003). Increasing entropy in the returns/volume system indicates a greater need for off-market information to reduce disorder. Entropy changes in the return--volume space could be a new variable in machine learning forecasting algorithms, reflecting system unpredictability or indicating demand for off-market information.

The analysis aims to determine mutual relations between entropy measures for different markets and quotation frequencies during periods causing economic shocks. We used entropy measures to compare changes in financial markets before, during, and after the COVID-19 pandemic and the start of the war in Ukraine. Beyond Shannon, Theil, and relative entropy, we use the Tsallis distribution to study memory effects for all returns and separately for positive and negative ones. A higher q-parameter indicates higher memory effects and system order. For two-dimensional analysis, we apply entropy-based Kullback-Leibler Divergence. Entropy-based analysis of returns and liquidity has not yet been applied to stock markets. This exploratory study investigates how these entropy measures behave and relate to each other, determining if they are good indicators of shock changes in financial markets caused by real processes.

We analyse two stock market indices: the US S&P 500 and the Polish WIG20, in four selected years: 2019 (pre-pandemic), 2020 (pandemic), 2021 (quasi post-pandemic), and 2022 (post-pandemic and during the war in Ukraine) using data of three different frequencies: minute, hour, and day. Selected periods and indices are treated as case studies. Our exploratory study aims to uncover patterns and relationships in data not well-defined in the literature. We predict that entropy measures can indicate and identify the impact of unexpected external shocks on financial markets. Our goal is to examine whether analysing financial market signals, such as returns and volume, collectively provides more insight than analysing them separately. We hypothesize that if there are no differences in the behaviour of these measures between stable and disruptive periods, and no differences in the frequency of quotations across different markets, further research would be unnecessary.

2. Entropy-based methods in economics and finance

Entropy is a concept of notable complexity, with various interpretations that have even permeated popular culture (Smith 2017; Ben-Naim 2019). Its transition from pure thermodynamics to probability and statistics occurred in the late 19th century when Boltzmann explained thermodynamics through the motion of molecules. This concept also served as the foundation of information theory (Natal et al. 2021). Economics, borrowing extensively from the realm of physics, has incorporated various forms of entropy to address different economic concerns. Broadly, for economists, entropy serves as an indicator of system disorder. Yet, it can also be interpreted as a measure of risk or uncertainty within financial models, an estimation of inefficiency in production theory, or a measure of goodness-of-fit in optimization methods. Economists often combine diverse approaches and use this concept as a metaphor for social processes (Jakimowicz 2020). Like physicists, economists strive to formulate a generalized framework to accommodate similar yet distinct concepts of entropy. In recent studies, Wolfram (2023) has unified thermodynamic, statistical, and information-based approaches using computational foundations. This attempt appears particularly promising, especially within social science, where computational methods, machine learning, and artificial intelligence are gaining popularity.

The concept of entropy in physics originated from observations that in steam engines, a significant portion of energy was lost due to friction and dissipation, rendering it unconvertible into useful work (see the historical perspective of the evolution of the concept in Wolfram 2023). According to the second law of thermodynamics, this energy can't just disappear. Clausius (1867) described this missing energy by entropy and presented the first mathematical definition of entropy.

$$S = \frac{1}{T}Q \tag{1}$$

where Q represents the quantity of heat absorbed by the system from its external reservoir, and T denotes the temperature of this reservoir (or the surroundings) at a specific moment in time.

Boltzmann (Wolfram 2023) formulated a different definition of entropy, being a statistical measure of the molecular disorder of the system. It has the following form:

$$S = k \ln W \tag{2}$$

where k is the Boltzmann constant, while W is the total number of microscopic states, corresponding to the macroscopic state of the system.

Equivalently in statistical physics, the Boltzmann-Gibbs entropy is used, given by the formula:

$$S_{BG} = -k \int \rho(x, t) \ln \rho(x, t)\, dx \tag{3}$$

where $\rho(x, t)$ is the probability density function of the stochastic variable x (generally time-dependent).

The concept of thermodynamics was adapted to economics by Fisher (1925), who translated the fundamental principles of physics into economic terminology. In this translation, material points (particles) were equated to economic entities (individuals), forces were replaced by the concept of marginal utility, and energy was paralleled with utility. This allowed for the transposition of the law of equilibrium from physics to economics. In this context, the equilibrium point, which in physics is determined by the maximization of net energy, in economics corresponds to the maximization of the profit function.

The notion of entropy first appeared in economics in 1971 when Georgescu-Roegen published *The Entropy Law and Economic Process*. This work laid the foundation for the theory of production by applying the second law of thermodynamics to economic considerations. In this view, entropy could be understood as an irretrievable loss of productive energy that cannot be transformed into useful work.

Entropy also measures the deviation of the analysed distribution from full concentration (minimum entropy) to full dispersion (maximum entropy). Full dispersion typically occurs with a uniform distribution, where the probabilities of all events are equal. In terms of predictability, the lower the entropy, the lower the uncertainty, and the higher the predictability. This approach links entropy to the efficient markets hypothesis: the greater the entropy, the greater the efficiency of financial markets (Rothenstein 2018).

In economics and finance, the concept of information entropy, as formulated by Shannon (1948), is extensively applied to analyse and quantify uncertainty. The Shannon entropy is expressed as follows:

$$H = -\sum_{n=1}^{N} p \ln p \tag{4}$$

where p is probability of point event and $N = 1, 2, ..., n$ is the number of events.

The maximum value of Shannon's entropy for uniform distribution (all events have equal probabilities) is $p = 1/n$ and takes a value $H_{max} = -n \cdot (1/n) \cdot \ln(1/n) = \ln(n)$. The minimum value of Shannon's entropy is equal to 0 and is available only when all events are concentrated in one point. If there are two equally likely events with $s = 1/2$, entropy is equal to 1. To find the missing gap between observed and potential entropy, it is applying the relative entropy given by the formula:

$$R = \frac{H}{H_{max}} = \frac{H}{\ln n} \tag{5}$$

where H is the measured entropy and $\ln n$ is the maximum entropy for n events.

Hence $R = 1$, when events are fully diversified and $R = 0$ if events are fully concentrated.

The additional measure is Theil's entropy (called also redundancy). It measures how much Theil's entropy deviates from the maximum Shannon entropy:

$$T_{Theil} = H_{max} - H \tag{6}$$

where H_{max} is Shannon's maximum entropy (for equal distribution) and H is Shannon's entropy for observed data.

In other words, Theil's entropy measures the gap between observed and maximum entropy.

The concept of entropy was applied to portfolio selection for the first time by Philippatos and Wilson (1972). They proposed a mean-entropy approach and compared it to traditional methods. They found consistent mean-entropy portfolios with the Markowitz full-covariance and the Sharpe single-index models. Entropy has been applied in option pricing. Gulko (1997) introduced the Entropy Pricing Theory (EPT), which can offer some similar valuation results equal to the Sharpe-Lintner capital asset pricing model and the Black-Scholes formula. The EPT was also applied to stock option pricing Gulko (1999) and bond option pricing Gulko (1999, 2002). Buchen and Kelly (1996) used the Principle of Maximum Entropy to estimate the distribution of an asset from a set of option prices. They showed that the maximum entropy distribution was able to fit a known probability density function accurately. Thanks to that, simulating option prices at different strike prices is possible. The maximum entropy method could be used also to estimate the implied correlations between different currency pairs (Krishnan, Nelken 2001) and to get the implied probability density and distribution from option prices (Guo 2001; Borwein, Choksi, Maréchal 2003).

Entropy is also strongly connected to Tsallis distribution – it maximizes Tsallis q-entropy, in a similar way as Gaussian distribution maximizes Boltzmann-Gibbs entropy. Thus, the Tsallis entropy is a generalization of Boltzmann-Gibbs entropy (Tsallis 1988). It estimates the fat-tailed probability distribution better than the Gaussian distribution. Fat-tailed distributions of price returns are well documented in empirical studies (Cont 2001; Bil, Grech, Zienowicz 2017). Fat tails are the extreme returns (large events) that may happen – they are more probable than predicted with the normal distribution (Gaussian distribution underestimates their probability, although it is often used in finance, i.e. Fama (1965) and Merton (1973)). They are usually measured with skewness (the third moment) and kurtosis (the fourth moment) as kurtosis is very sensitive to outliers (Kopczewska 2014).

According to the law of equilibrium, Boltzmann-Gibbs entropy is maximized if the probability density function of its states $\rho(x)$ is Gaussian. As the Gauss distribution does not fit to fat-tailed probability distributions, in the case of financial market data, which have fat-tailed distributions, more suitable is the aforementioned Tsallis formalism. Tsallis et al. (1995) proposed a generalization of Boltzmann-Gibbs entropy, called Tsallis q-entropy S_q, given by the formula:

$$S_q = k\frac{1}{q-1}\left(1 - \int\rho\left(x,t\right)^q dx\right) \tag{7}$$

Additionally, Tsallis introduced a distribution of probabilities called q-normal or Tsallis distribution:

$$p\left(x\right) \approx N_q\left(1 + B_q x^2\left(q-1\right)\right)^{\frac{1}{1-q}} \tag{8}$$

where q is the single continuous arbitrary real parameter, while B_q and N_q are as follows:

$$B_q = \left[\left(5-3q\right)\sigma^2\right]^{-1} \tag{9}$$

and

$$N_q = \frac{\Gamma\left(\frac{1}{q-1}\right)}{\Gamma\left(\frac{3-q}{2q-2}\right)}\sqrt{\frac{q-1}{\pi}B_q} \tag{10}$$

while Γ is the Gamma-Euler function:

$$\Gamma(x)=\int_0^{+\infty} t^{x-1}e^{-t}\,dt \tag{11}$$

Tsallis distribution maximizes Tsallis q-entropy, in a similar way as Gaussian distribution maximizes Boltzmann-Gibbs entropy. The Tsallis entropy is a generalization of Boltzmann-Gibbs entropy, therefore, is reduced in the limit $q \to 1$ to the classical Boltzmann-Gibbs form, the same as Tsallis distribution in the limit $q \to 1$ is reduced to Gaussian distribution. Fitting q-normal distribution to normalized returns consists in changing q-parameter. To apply this, the mean squared displacement method (MSD) is used:

$$MSD \equiv \left(x(d)-x(q)\right)^2=\frac{1}{N}\sum_{n=1}^{N}\left(x_n(d)-x_n(q)\right)^2 \tag{12}$$

where N is the number of particles to be averaged, $x(d) = x_n(d)$ is the n-th real data, $x(q) = x_n(q)$ is the n-th estimated data from the Tsallis distribution.

The goal is to get a q that minimizes MSD. To find the asymmetry in price returns, the fit can be done twice: first, for all returns, which gives the best-fit q value for all returns independently of their sign. Then the separate fits are performed for positive and negative returns independently. The q-parameter is a determinant of memory level in the system. A higher q-parameter means higher memory effects. Along with increased memory effects, the level of order in the system increased.

This Tsallis formalism is very useful in studying the behaviour of the stock exchange, but is not popular in economic articles. Many interesting results can be obtained with the Tsallis distribution. In Bil, Grech and Zienowicz (2017), the authors show some of their possible applications. Firstly, they show that the q-normal distribution is much better fitted to the real distribution of price returns. Additionally, they discover the difference between the behaviour of negative and positive returns. Furthermore, this behaviour depends on the market and frequency of data sampling. The emerging markets turn out to obtain an equilibrium state faster than mature markets. With the decreasing frequency of data sampling distribution of price, returns become close to random ($q \to 1$). Interestingly, positive returns become random usually faster than negative. More interesting conclusions can be made by analysing q-parameters, especially their asymmetry δ_q. Information on both $\delta_q = q^+ - q^-$ and q values are very useful to get information about the state of the complex financial system – especially about different behaviour of asymmetry index δ_q between interday (1 day – 4 day time lags) and intraday (1 min – 60 min time lags) trading. Interday price returns feature a higher asymmetry ratio than in the case of intraday returns. The authors suggest that this effect is caused by the different types of investors who participate in intraday and interday trading. The first type of traders are mostly institutional investors, who trade usually on a shorter time scale using econometric models and numerical applications. The second type are individual traders, who use more traditional technical analysis.

The Kullback-Leibler Divergence (KLD) (Kullback, Leibler 1951) is a two-dimensional entropy-based measure. It is one of the most important entropy optimization principles and a useful method to find the difference between two probability distributions. It is a slight modification of the formula for the Shannon entropy of variable X, given in continuous case by the formula:

$$H(X) = \int f(x) \ln \frac{1}{f(x)} dx \tag{13}$$

where $H(X)$ can be viewed as the expectation of the quantity $\ln[1/(f(x)]$, that can be called a loss.

In decision theory, a risk is the expectation of a loss, so that $H(X)$ can be called a risk. When there are two distributions g and f for a random variable X, we can define the cross-entropy of g relative to f. For a discrete variable taking values x_j, $j = 1,\ldots, m$ and denoting $p_j = f(x_j)$ and $q_j = g(x_j)$, we have:

$$CE(g \mid p) = \sum_{i=1}^{N} p_j \ln \frac{1}{q_j} \tag{14}$$

It can be viewed as the expectation of a loss indicator $\ln[1/(g(x)]$ under the distribution f. It can be decomposed in:

$$CE(g) = \sum_{j=1}^{m} p_j \ln \frac{p_j}{q_j} + \sum_{j=1}^{m} p_j \ln \frac{1}{p_j} \tag{15}$$

The first part of this equation is called the relative entropy or Kullback-Leibler Divergence of g relative to f:

$$D_{KL}(g \mid p) = \sum_{j=1}^{m} p_j \ln \frac{p_j}{q_j} \tag{16}$$

We can tell that if the observations come from f, the risk associated with g is:

$$CE(g \mid f) = D_{KL}(g \mid f) + H(f) \tag{17}$$

that is, this is the sum of the Kullback-Leibler Divergence, and the entropy of f (the risk already associated to f).

Kullback-Leibler Divergence is sometimes called Kullback-Leibler distance. Indeed, it is a measure of how far g is from f and it has some of the properties of a distance: for all f, g we have $KL(g|f) \geq 0$; we also have $KL(f|f) = 0$, and conversely if $KL(g|f) = 0$ then $f = g$ nearly everywhere. However, it is not a distance because it is not symmetric and does not satisfy the triangular inequality. Usually, p_j represents the data, the observations, or a measured probability distribution. Distribution q_j represents instead a theory, a model. Kullback-Leibler Divergence is not often used in economic and financial articles, but Kim and Sayema (2017) effectively use KLD to predict stock market movement. The solution, how to assess two distributions and the direct divergence between them, is a two-dimension Kullback-Leibler Divergence, which can be expressed in two ways, as follows:

$$2D_{KL}^{i}\left(g \mid p\right)=\sum_{j=1}^{m} p_{ji}\ln\frac{p_{ji}}{q_j} \tag{18}$$

where q_j stands for the probability of point event in interval j of f-distribution in the g-distribution and p_{ji} is defined as the share of the probability of a point event in a given interval of f-distribution in a given g-distribution with reference to the full intervals probability.

KLD is thus the sum by m intervals (j = 1, 2,..., m), or:

$$2D_{KL}^{i}\left(g \mid p\right)=\sum_{j=1}^{m} p_{ji}\ln\frac{p_{ji}}{q_j} \tag{19}$$

where q_i is the expected value of point event in interval j of g-distribution in the f-distribution analysed, and p_{ji} is the share of the probability of a point event in a given interval of f-distribution in a given interval of q-distribution regarding full intervals probability.

The generalized Kullback-Leibler Divergence in Tsallis statistics (TKLD) (Devi 2018) is simply given by change ln(x) on q-logarithm $\ln_q(x)$ in Kullback-Leibler Divergence KLD, consequently one gets:

$$D_q(g \mid f)=\sum_{j=1}^{m} p_j \ln_q\frac{p_j}{q_j} \tag{20}$$

where q-logarithm $\ln_q(x)$ is:

$$\ln_q(x)=\frac{x^{q-1}-1}{q(q-1)} \tag{21}$$

Additionally, Huang, Yong and Hong (2016) proposed to modify the q-logarithm by giving an additional q that has been inserted into the denominator to keep well-defined TKL Divergence both positive and non-increasing under transformation for all $q \in R\{0, 1\}$. Finally one gets:

$$D_q(g \mid f)=\sum_{j=1}^{m} p_j \widehat{\ln}_q\frac{p_j}{q_j} \tag{22}$$

where:

$$\widehat{\ln}_q x=\frac{x^{1-q}-1}{q(q-1)} \tag{23}$$

The development of the theoretical approach to entropy has allowed information entropy and mutual information to gain importance in contemporary economic theory and research. We can identify three primary domains where their influence on economics and finance is expanding: (i) as measures of uncertainty employed in unorthodox economics, (ii) as a new tool for identifying relationships among variables, including cause and effect relationships, and (iii) as measures of goodness-of-fit and instruments for variable selection in statistics and machine learning.

Applying entropy-based measures in econophysics revives the methodological debate regarding the distinction between risk and uncertainty within economic theory and research. While these concepts should not be used interchangeably, academic textbooks and scholarly articles often fail to clearly separate them (King, Kay 2020). Many uncertain situations have been simplified into measurable categories, transforming them into risk scenarios with predefined probability distributions (Mandelbrot 1997). In financial economics, decision-making under risk has been reduced to mean-variance criteria. Economists predominantly place the concept of uncertainty in the epistemological realm. This has led to the decline of uncertainty theories proposed by Knight, Keynes, and Hayek in contemporary economic discussions due to their immeasurability. Entropy-based measures offer a new way to quantify uncertainty, increasing methodological pluralism in economics and restoring the importance of these uncertainty theories (Schinckus 2009). Despite methodological disputes, researchers in economics and finance are increasingly using entropy measures as alternatives or additions to standard risk measures like volatility or Value at Risk. Entropy is better at explaining non-linear dynamics and tail risks, which are crucial in financial risk management (Ahmadi-Javid 2012).

Time plays a crucial role in defining both entropy changes and causality. The second law of thermodynamics posits that an increase in entropy dictates the direction of time flow. Similarly, causality, which concerns the relationship between cause and effect, inherently requires a temporal direction (Riek 2020). These parallels have made mutual information measures useful for enhancing standard procedures of analysing causality in the Granger-Sims sense, where “the past and present may cause the future, but the future cannot cause the past” (Granger 1980).

Schreiber’s (2000) transfer entropy is a statistical tool that measures the flow of information in a time-specific direction between random variables (Marschinski, Kantz 2002). This allows for the estimation of information transfer that is not symmetrical and focuses on the direction from cause to effect (Nichols, Bucholtz, Michalowicz 2013). Compared to the Granger-Sims approach, transfer entropy does not rely on any predefined econometric model and can detect nonlinear relations between variables. Quantifying causality using entropy has spurred researchers to conduct studies that complement causal inference between financial time series (Syczewska, Struzik 2015). They have investigated the flow of information asymmetry between individual financial markets, assessed the impact of stock indices on individual quotations, and analysed the relationship between returns and trading volume for individual financial assets (Keskin, Aste 2020; Marschinski, Kantz 2002). The rise of high-frequency textual data from social media platforms enables real-time tracking of market participants’ sentiments and expectations. Recent studies explore how this information influences investor behaviour and decision-making, focusing on the flow of information and causal quantification using transfer entropy (Yao, Li 2020).

Contrary to mainstream economics, which regards entropy as a supplementary concept in information theory, its importance is growing in statistics and machine learning. Entropy measures serve two primary purposes in these fields: as goodness-of-fit measures of models and as explanatory variables (features) in models. Machine learning increasingly uses information entropy’s core attribute – a measure of disorder – to enhance a model’s predictive power. High entropy indicates low information gain from using the model, whereas low entropy points to high information gain. Information gain can be seen as the quantity of useful knowledge in a system. Entropy minimization can be an objective function in optimization algorithms or support feature selection and dimensionality reduction in models, specifying the variables that most significantly impact the target variable. It is an alternative to the LASSO (Least Absolute Shrinkage and Selection Operator) method (AlMomani, Sun, Bollt 2020).

Using entropy to measure risk and quantify causality enables these measures to describe the current state of financial markets effectively. Consequently, they are becoming more prevalent as features (variables) in machine learning models. Information entropy measures, calculated over specific periods, offer an alternative to traditional volatility measures. Mutual information and transfer entropy measures can determine which market, asset, or index provides more predictive insight into the target variable (Kim et al. 2020).

Our research closely relates to the development of using entropy measures to analyse markets, aiming primarily to identify a singular measure that can distinctly mark periods of varying stability and predictability in the financial market. Firstly, we observe that investors prioritize analysing information derived from returns. When the efficient market hypothesis does not hold, volume information can become a second component, indicating behavioural factors contributing to market inefficiency. Our study views the market as a system emitting two signals without assigning primary importance to either. This dual-signal approach aligns with theoretical discussions on market equilibrium mechanisms. Bray (1981) demonstrated that an artificial market's behaviour is best described through two signals: one for asset returns and the other for random endowments spread among agents, with the effects of these signals being inseparable. The analysis of changes in a two--dimensional information system can be reduced to an analysis of changes in the joint entropy of a coupled information system of return and volume (R, V), where $P_R(r)$ is the PDF of the random variable R (returns), $P_{R,V}$ is the PDF of the random variable v (volume), and $H(R, V)$ is the joint entropy between R and V, it is characterized by the following (Madiman 2008):

$$H(R,V) = -\sum_{r\in R}\sum_{v\in V} P_{r,v}(r,v)\log P_{r,v}(r,v) \tag{24}$$

Consequently, in a case where variables R and V are independent, the joint entropy of the system is accurately represented as a simple sum of their individual entropies: $H(R, V) = H(R) + H(V)$. When the joint entropy $H(X, Y)$ is less than the sum of the individual entropies $H(X) + H(Y)$, this denotes the presence of mutual information $I(R, V)$, which is quantified as the difference between the sum of individual entropies and the joint entropy of the system: $I(R, V) = H(R) + H(V) - H(R, V)$. Unlike correlation, which only accounts for linear relationships, mutual information encompasses linear and nonlinear dependencies. The presence of non-zero mutual information signals a need for further investigation into the characteristics of these dependencies using KD divergent measures.

This type of research has not yet been carried out in finance. Multivariate and multiscale entropy measures are typically used to search for dependencies between time series of prices (Giannerini, Goracci 2023). However, advances in techniques for decomposing information in multidimensional systems in engineering and biology (Faes et al. 2016) are expected to be rapidly applied to the analysis of financial markets.

Our primary objective is to explore the relationships between returns and trading volume, as expressed through entropy measures, and to assess the utility of considering the return/volume relationship as a joint information system. To this end, we compare changes in Shannon entropy measures calculated independently for returns and volume over selected periods. Using relative entropy measures, our methodological approach involves calculating and analysing joint information measures, mutual information, and changes in the two-dimensional distributions of returns and volume. Modified measures such as Kullback-Leibler (KL) and Topsoe's KL Divergence (TKLD) are employed to estimate the difference in information represented by these two-dimensional distributions.

Previous studies have extensively documented the impact of COVID-19 on financial market returns and liquidity. However, research is sparse on the concurrent behaviour of these two parameters over time. Our analysis aims to ascertain the degree of disorder in financial markets during the pandemic and its subsequent evolution. A crucial aspect of this research is determining if significant variations in entropy measures exist across these four periods, each characterized by unique external influences. If no notable differences are identified, it would suggest that the practical application of these measures, in both theoretical constructs and predictive models, is limited, given the distinct nature of the chosen periods.

3. Statistics of the financial market

In the initial versions of our work, we selected several crisis periods and various sample auction and stock index quotations, along with gold and currency rates, to isolate a typical relationship between the behaviour of entropy measures during crisis periods. However, this approach proved inadequate for entropy measures. Research on causal relationships based on entropy measures indicates their ability to isolate partial dependencies. Unlike linear models such as Granger-Sims causality, entropy measures can indicate non-linear dependencies that may only become apparent after certain thresholds of the variables are exceeded.

From the beginning, we could not count on finding a general pattern of behaviour of entropy measures in periods of turmoil, so we limited ourselves to periods with the most contrasting characteristics of the macroeconomic environment and a limited number of quotations from markets with significant differences. This exploratory study focuses on various measures of information entropy to investigate their behaviour across four distinct periods, each characterized by unique external influences: during and after two global shocks – the COVID-19 pandemic (from March 2020 onward) and the war in Ukraine (from March 2022 onward). These two periods, the COVID-19 pandemic and the war in Ukraine, are treated as natural experiments, contrasted with a control period before the pandemic and one after the shock period.

We examined the price returns and trading volume of two stock indices from very different markets: the WIG20 index from the Warsaw Stock Exchange in Poland and the Standard & Poor's 500 (S&P 500) index listed on the New York Stock Exchange. We analysed two distinctly characterized markets: an emerging market with relatively small capitalization and low liquidity and a developed market characterized by large capitalization and significantly higher liquidity. The analysis was carried out at the market level rather than at the level of individual assets since market indices, as aggregates, demonstrate less sensitivity to external information than individual assets. Choosing a specific asset for a case study would entail a problem of result representativeness.

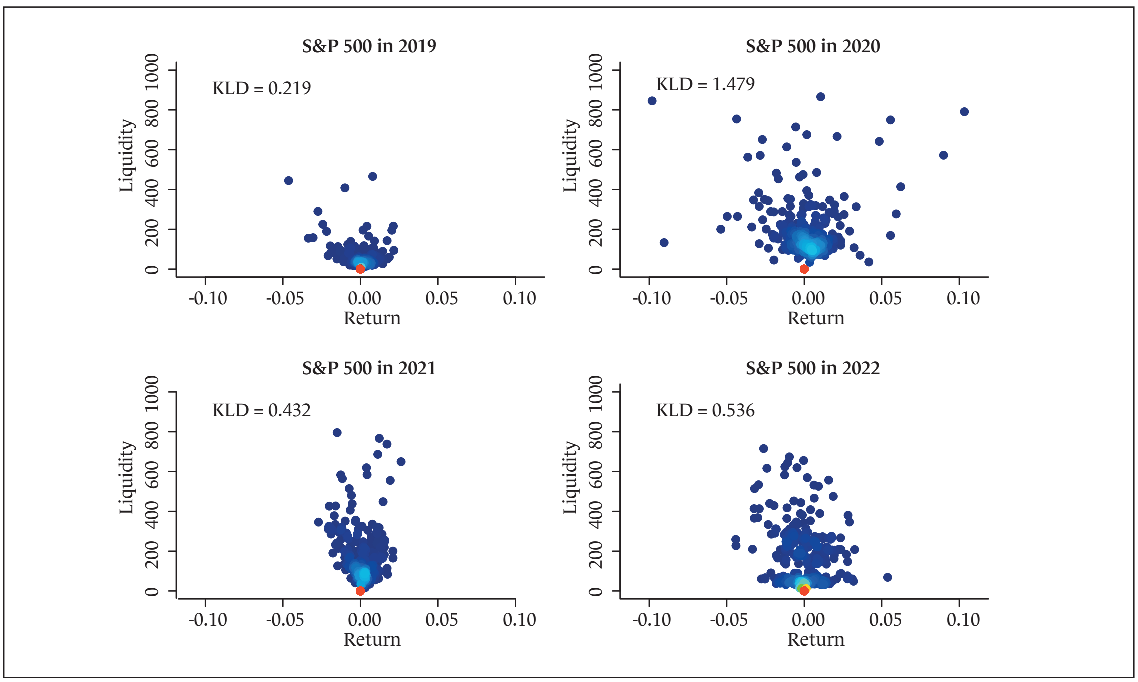
The statistics of these market indexes were taken from www.dukascopy.com. We analysed four annual periods: one year before the COVID-19 pandemic 1 March 2019 – 28 February 2020 (this period is hereinafter referred to as 2019), during the first year of the pandemic 1 March 2020 – 28 February 2021 (referred to as 2020), quasi-after pandemic period 1 March 2021 – 28 February 2022 (referred to as 2021), and after the pandemic and during the war in Ukraine 1 March 2022 – 28 February 2023 (referred to as 2022). The periods we chose are each one year long and start right after two external shocks: the implementation of COVID-19 restrictions and the outbreak of the war in Ukraine.

We analyse the data of three frequencies: day, hour and minute. Based on the previous research we conclude that the frequency of data has crucial impact on estimation entropy measures (Bil, Grech, Podhajska 2016). We use relative returns $(P_{last} - P_{first})/P_{first}$, where P_{first} and P_{last} are respectively the first and last price of the period. They were calculated from period to period (e.g. from hour to hour, day to day, called time lag) and for the whole sample year. We used no methods to adjust the time series, preserving the raw information produced by the system for further analysis.

Initially, we analysed the basic parameters of the time series across all annual periods. Table 1 shows the average returns (calculated as the mean) and volatility (measured as the standard deviation) of these periodical returns. As anticipated, we observed a time-delayed increase in volatility. In 2020, due to COVID-19, volatility at least doubled compared to 2019. It then dropped to about 1.2–1.4 times the 2019 level, before rising again to remain at 1.6–1.75 times the 2019 volatility. This pattern suggests that stock market fluctuations were more pronounced during global shocks, with COVID-19 having a more significant impact than the war in Ukraine. Meanwhile, the average periodic returns hovered around zero, showing no clear trend.

Figure 1

Scatterplots of returns and liquidity of daily data – S&P 500



Note:

Figure 1 offers a clear visualization of this pattern. It is a two-dimensional scatterplot showing the relationship between returns (on the x-axis) and liquidity (trade volume, on the y-axis) for daily S&P 500 data over consecutive years. The plot clearly reveals structural changes between pandemic and non-pandemic years. While the periodic returns remain similar, the range of returns (x) and liquidity (y) vary significantly between these markets.

Subsequently, we checked the difference between the annual returns over the analysed periods (2019, 2020, 2021, 2022) – it was -25%, 5%, 1%, -5% for WIG20 and 6%, 24%, 11%, -8% for S&P 500. It shows an increase in returns during the pandemic and a significant drop during the war in Ukraine in both markets.

Following Bil, Grech, and Zienowicz (2017), we suspect that price returns also have fat tails. Therefore, we calculated skewness and kurtosis for every security and all periods (Table 2). The results of skewness show that before the shocks all the returns distributions were the most left-skewed, while the COVID-19 shock made them more right-skewed, with the peak of right-skewness in 2021, after the pandemic. Kurtosis for all time frequencies decreased as the time lag grew, which is evidence that fat-tails are mostly the case of high-frequency data. The highest kurtosis was in 2020, which suggests that serious shocks such as COVID-19 change the returns distributions into fat-tailed.

The volume of transactions is evidence of the market liquidity (Table 3). During the 2020 pandemic, the average transaction volume surged by 3–4 times before dropping to 80–90% of its 2020 peak on the US market. This confirms that during the pandemic market activity was greater than before and this change became persistent. In Poland, the shock linked to the war in Ukraine increased market liquidity, while COVID-19 did not boost interest in the stock market as it did in the US. Intuitively, investors are more nervous in turbulent times, which results in greater fluctuations in the market. When the financial market is dominated by institutional players (as in Poland), emotional behaviours are less visible.

We also checked the memory effects in data using a q-parameter from the Tsallis distribution (Table 4). We used this q-parameter to fit the Tsallis distribution to the empirical distribution of price returns of analysed securities. The fit was done twice: first, for all returns, which gives the best-fit q value for all returns independently on their sign (q); and secondly, for positive (q^+) and negative (q^-) returns independently. Additionally, we calculated the relative asymmetry ratio of the Tsallis distribution as $|(q^+) - (q^-)|/q = |\delta_q|/q$.

A higher q parameter indicates stronger memory effects in price returns. As shown in Table 4, the q parameter decreases with a growing time lag Δt – its values were between 1.52 and 1.20. This indicates that memory in time series is gradually being lost with a growing time lag. The q-parameter for daily data is much smaller than for 1 sec or 1 min data. In general, the post-pandemic data (2021, 2022) have lower memory (lower q-parameter) than the pandemic data. The relative asymmetry ratio of the Tsallis distribution was the highest in 2022, which suggests one-sided frictions during the war in Ukraine – on both markets memory effects of daily data were significantly stronger (23–30%) for losses than for profits. However, the memory effects do not have persistent regularities. Frequency, a moment of time and market matter for those parameters.

We also tested the disorder of the market using Shannon entropy, applied to returns and volume separately. Theoretically (Table 5), higher entropy (of returns) reflects the higher uncertainty or risk of financial markets and higher efficiency. The lower the entropy, the lower the risk and higher predictability of the market, and thus lower efficiency (thus inefficiency). In the case of liquidity, the relation is similar – the higher the entropy, the deeper the market.

When analysing indicators of market disorder (Table 6) one should take the reference maximum entropy for all measurements equal 1.61 ($H_{max} = -5 \cdot (0.2 \cdot \log(0.2))$), which is for five intervals assuming a uniform distribution. Empirical entropy values for returns and volume were calculated using five intervals. One can observe a few regularities.

First, the longer the time lag, the higher the entropy. Thus daily data show the market is much more efficient than minute data. That is true for both indices, returns and volume in all years presented here. However, markets differ in the degree of (in)efficiency due to data frequency. On the Polish market, the entropy of daily returns is approx. 50 times higher than that of hourly returns. On the US market, this difference is halved, around 25 times. This means that the US high-frequency market has the same efficiency as the Polish one, as entropies are similar. However, with daily data the Polish market is much more efficient than the US market. The trade volume of WIG shows high entropies, with daily data close to the maximum one which is evidence of a deep, liquid market. The US results look very similar to the Polish ones.

Secondly, the impact of the COVID-19 shock was evident in both markets. The entropy of returns increased approximately 2 to 18 times in 2020, then dropped by 20–90% the following year. The shock from the Ukraine war was less pronounced, leading to an increase of 1 to 10 times entropy. The behaviour of trade volume differed: on the Polish market, entropy remained relatively stable for hourly and daily volumes, with fluctuations of up to 20% observed in daily data. However, trade volume grew by 25% in 2022 during the Ukraine war, indicating market growth (as confirmed by a significant increase in trade volume, see Table 3). The US market responded differently: the COVID-19 shock increased volume entropy by approximately 2 to 4 times, suggesting a deeper market, followed by a drop to a new stable level. Since the pandemic, the S&P 500 has attracted significantly more investors than before.

4. Two-dimensional relations return/liquidity on financial market

Two-dimensional entropy-based analysis of stock markets in general involves KLD statistics, which compares two different time series. They are transformed to the form of a two-dimensional cross-table, which counts in intervals the number of observations which fulfil both criteria (Kopczewska et al. 2017). The number of intervals must be the same for each vector. The KLD measure is sensitive to the number of intervals. Therefore, few outliers were rescaled to match the assumed intervals. KLD has in general higher values when fewer intervals are included. KLD takes the value 0 when two analysed vectors are the same and KLD > 0 in the case of differences. The higher the KLD, the higher the dissimilarity.

For the return/liquidity setting, we assume that both vectors (returns and liquidity) are connected by the day they happened and are of equal length. We assumed five intervals, both for returns and liquidity (however with different thresholds for both features) – this covers all observations from all analysed periods. For daily WIG20 data, the intervals were from -5% to 5% for returns and from 0 to 3 million. For S&P 500 daily data intervals were from -10% to +11% for returns and from 0 to 1 billion for volume.

Table 7 presents KLD statistics between returns and liquidity calculated for each year. In fact, it answers the question of dispersion of data over derived intervals (e.g. if observations are concentrated in a few selected intervals or more combinations of returns/liquidity appeared over the whole span range). The analysed two-dimensional distributions of return/liquidity were visualised in Figure 1. In the case of the S&P500 one can observe a significant change of the structure in this relation in the pandemic year 2020 – higher dispersion of data (Figure 1) is confirmed with high value of KLD (Table 6). Other years (2019, 2021, 2022) have a more similar structure of returns and liquidity. In Poland, structural changes in daily data were longer and peaked in 2021. The Polish market came back to a similar pre-COVID structure, while the American market stayed more dispersed.

KLD can also compare market structures between years. The analysis in the previous step generated a cross-table for return and liquidity. Those two-dimensional tables can be vectorized and then become an input to KLD. Figure 2 presents the visualization of the KLD matrix for the daily S&P 500, which compares two-dimensional settings in two selected years. This is to answer the question whether market structures return/liquidity in the analysed two years were similar. The diagonal of this matrix compares the same data (the year 2019 with 2019, etc.), thus KLD = 0. The highest KLD is for the comparison of the years 2019 (pre-pandemic) and 2020 (pandemic) (KLD = 0.01765, dark blue) – this is evidence that the market structure of return/liquidity changed significantly. KLD between the years 2020/2021 (KLD = 0.00063, gray) and 2021/2022 (KLD = 0.00077, gray) are already low and show only slight differences between those years, which can be explained as quickly coming back to the regular "non-shock" state. However, KLD between the years 2019 and 2022 (KLD = 0.01074, light blue) is evidence that the post-pandemic equilibrium is not the same as the pre-pandemic one. As seen in Figure 1, even if returns stayed in similar intervals, the volume of trade changed significantly (it approximately tripled). The very different pandemic (2020) KLD can be explained by a much higher dispersion of returns/liquidity values than in stable years. This is what entropy confirms – the transition from one equilibrium state to a new equilibrium state.

Figure 2

Kullback-Leibler Divergence (KLD) across years for two-dimensional settings: return/liquidity; S&P 500 daily

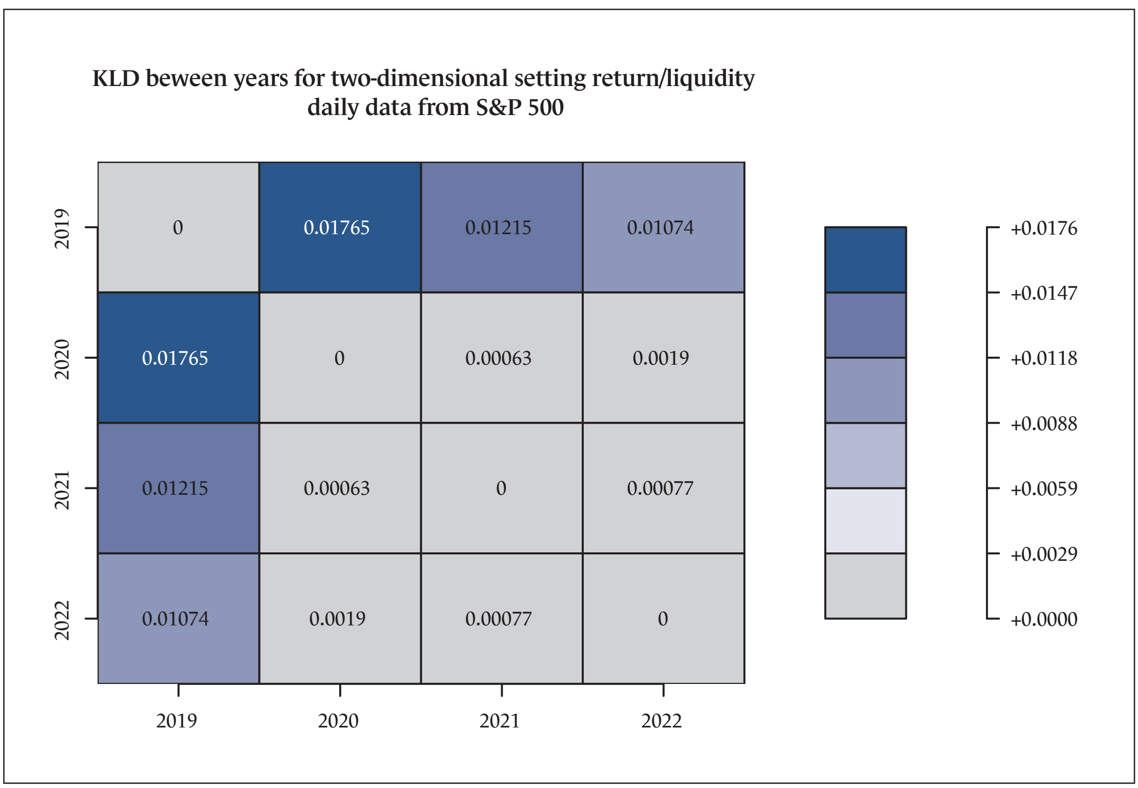


Table 8 reports KLD matrices between years for return/liquidity setting for different indices and data frequency. All annual cross-tables were built using five intervals. One can see that on the US market independently of data frequency, the impact of the pandemic shock was visible, while the war in Ukraine had almost no effect. On the Polish market, the impact of the pandemic shock causing structural changes in the market appeared too and was also strong in 2021, while the war in Ukraine in 2022 had a separate strong influence.

The results of traditional KLD can be extended to Kullback-Leibler Divergence in the Tsallis statistics (Table 9), following (Devi, Sandhya 2018), University Library of Munich, Germany. The results of generalization in Tsallis statistics kept the same structure as traditional KLD. They suggest a deeper pandemic shock in 2020 on the S&P 500 and WIG20, and a temporal shift of this shock in 2021. The shock of the war in Ukraine is also visible in both markets.

5. Concluding remarks

This article is an exploratory analysis of the use of a two-dimensional entropy measure in the return/volume space in financial markets. The first part of the article presents a theoretical approach to the possibilities of using derivation of a two-dimensional entropy measure. In the second part, as a case study, two selected stock indices were analysed for three periods and for three quotation frequencies. The value added by this paper is a two-dimensional approach to analysing returns and liquidity trading jointly. This approach is still rare in economic and financial literature.

From a theoretical standpoint, the significance of entropy measures in modelling market instability is gaining traction. This trend is attributed to the convergence of physics and economics, as well as the universal attributes of entropy as a measure. Despite its extensive history, the concept of entropy continues to acquire novel interpretations and applications, particularly within the domain of information theory (Wolfram 2023). Financial markets, characterized as systems that absorb and generate information, present a conducive environment for applying these innovative approaches. Notably, mutual and transfer entropy have become increasingly pivotal, as they facilitate the tracking of information flow and the delineation of causal relationships (Yao, Li 2020). There is an observable shift from one-dimensional entropy measures towards high-dimensional counterparts (Giannerini, Goracci 2023). Applying entropy measures to the returns/volume space aligns with this emerging research trajectory.

In the empirical part, we present evidence that financial markets react to local and global shocks, which is reflected in structure of returns/volume entropy measures. We used Kullback-Leibler Divergence (KLD) in its traditional form and in the Tsallis statistics to inspect the return/liquidity structure of US and Polish financial markets, using the S&P 500 and WIG20 indices. We detected a significant structural change in the S&P 500 caused by COVID-19 using KLD and showed that in the post-pandemic periods the market did not return to the same equilibrium as in the pre-pandemic times. Additionally, with Tsallis entropy, we demonstrated asymmetry in market memory, indicating that profits and losses do not mirror each other symmetrically. This is connected to asymmetry in market returns detected with skewness and kurtosis. Many investors are aware that financial market models work well when assumptions of efficiency and the deep market hold. As we show, this is not always the case, due to external shocks and measurement issues – daily data evidence a much more efficient market than minute data. These permanent and/or temporary shifts significantly impact the quality of predictions, model calibration, and the way markets are modelled. Having a reliable toolbox to capture these structural changes and determine whether they are permanent or temporary is a crucial issue in long-term financial market modelling. One can question why these phenomena appear. This opens another stream of studies on behavioural and cognitive traps, widely discussed in the literature (e.g. Merton 1987), as well as the rationality of investors treated as economic agents. But the measure of the impact of these disturbances on the behaviour of the market as an information system can be two-dimensional entropy in the return/volume space.

There are also new unsolved issues related to this study. First, how much can we rely on historical data for predictions? In the case of a regime change in financial markets, there is no basis for using historical data to forecast current changes. Second, are the changes we observe temporary or permanent, and will markets return to their previous levels observed before the shock? Third, what frequency of data should be considered standard for informative comparisons? High-frequency trading algorithms exploit inefficiencies in the data. Any information about the state of the market will make it more efficient (at least in a weak sense). As we show, entropy measures are frequency- and data-range-dependent, requiring a standard reporting method to guarantee the availability of comparisons. Finally, the problem with entropy measures is their lack of easy translation into commonly used financial models. Primarily, the results obtained from estimating entropy measures are not interpretable. This is not recommendable in financial models, especially those created for regulatory authorities. While entropy is considered a superior measure of volatility in portfolio analysis (Philippatos, Wilson 1972) and as a predictor (Maasoumi, Racine 2002), it is unlikely to replace traditional measures like volatility expressed through standard deviation in financial models. This resistance to change can be attributed to the industry's widespread acceptance and historical precedence of standard deviation, along with its simplicity and ease of interpretation. The situation becomes even more complicated with mutual entropy measures, which can capture partial and nonlinear relationships between the analysed variables. For a given financial asset, the return/volume relationship may have a defined direction or strength after exceeding a particular threshold value of the variables. This relationship can also be asymmetric, with different strengths acting during rises and falls. For example, deriving entropy measures from the Tsallis distribution allows asymmetry in returns to be estimated, where the asset's volatility (measured by entropy) reacts differently to positive and negative returns. Mutual entropy and its components can capture asymmetry in the relationships between returns and volume. The measures can capture idiosyncratic asymmetry, which means these relationships are not solely the result of market factors but also events specific to a particular company or its characteristics.

Entropy measures convey information different from, for example, the VIX index. The information is more complex, but due to their conditional and idiosyncratic nature, the mutual relationship between their behaviour and their reaction to external shocks will be impossible to interpret and establish as a definite relationship. They will only be significant for forecasting/portfolio models that operate as black boxes, making them very interesting features for algorithmic trading but not for traditional financial models.

The results included in this article represent only a portion of the analysis conducted and the estimated entropy measures. To provide a more comprehensive view of the changes in the return/volume relationship and help readers better understand the concept of entropy, an online application is available on Shiny and R-CRAN at https://microeconomics.shinyapps.io/entropy/.

References

Aaronson S., Carroll S.M., Ouellette L. (2014), *Quantifying the rise and fall of complexity in closed systems: the coffee automaton*, arXiv:1405.6903.

Ahmadi-Javid A. (2012), Entropic value-at-risk: a new coherent risk measure, *Journal of Optimization Theory and Applications*, 155(3), 1105–1123, DOI: 10.1007/s10957-011-9968-2.

Addy W.A., Ajayi-Nifise A.O., Bello B.G., Tula S.T., Odeyemi O., Falaiye T. (2024), Machine learning in financial markets: a critical review of algorithmic trading and risk management, *International Journal of Science and Research Archive*, 11(1), 1853–1862.

AlMomani A.A.R., Sun J., Bollt E. (2020), How entropic regression beats the outliers problem in nonlinear system identification, *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 30(1), 013108, DOI: 10.1063/1.5129775.

Ben-Naim A. (2019), Entropy and information theory: uses and misuses, *Entropy*, 21(12), 1170, DOI: 10.3390/e21121170.

Bentes S.R., Menezes R. (2012), Entropy: a new measure of stock market volatility?, *Journal of Physics: Conference Series*, 394(1), 012033.

Bil Ł., Grech D., Podhajska E. (2016), Methods of non-extensive statistical physics in analysis of price returns on Polish stock market, *Acta Physica Polonica A*, 129(5), 986–992.

Bil Ł., Grech D., Zienowicz M. (2017), Asymmetry of price returns – analysis and perspectives from a non-extensive statistical physics point of view, *Plos One*, 12(11), e0188541, DOI: 10.1371/journal.pone.0188541.

Borwein J., Choksi R., Marechal P. (2003), Probability distributions of assets inferred from option prices via the principle of maximum entropy, *SIAM Journal on Optimization*, 14(2), 464–478.

Bray M. (1981), Futures trading, rational expectations, and market efficiency, *Econometrica*, 49(3), 575–596.

Buchen P.W., Kelly M. (1996), The maximum entropy distribution of an asset inferred from option prices, *Journal of Financial and Quantitative Analysis*, 31(1), 143–159.

Clausius R. (1867), *The Mechanical Theory of Heat: With Its Applications to the Steam-Engine and to the Physical Properties of Bodies*, John van Voorst.

Cont R. (2001), Empirical properties of asset returns: stylized facts and statistical issues, *Quantitative Finance,* 1(2), 223–236, DOI: 10.1088/1469-7688/1/2/304.

Devi S. (2018), *Financial portfolios based on Tsallis relative entropy as the risk measure*, MPRA Paper, 91614, University Library of Munich.

Dinga E., Oprean-Stan C., Tănăsescu C.R., Brătian V., Ionescu G.M. (2021), Entropy-based behavioural efficiency of the financial market, *Entropy*, 23(11), 1396, https://doi.org/10.3390/e23111396.

Faes L., Marinazzo D., Nollo G., Porta A. (2016), An information-theoretic framework to map the spatiotemporal dynamics of the scalp electroencephalogram, *IEEE Transactions on Biomedical Engineering*, 63(12), 2488–2496.

Fama E.F. (1965), The behavior of stock-market prices, *Journal of Business*, 38(1), 34–105, DOI: 10.1086/294743.

Fama E.F. (1970), Efficient capital markets: a review of theory and empirical work, *The Journal of Finance*, 25(2), 383–417.

Fisher I. (1925), *Mathematical Investigations in the Theory of Value and Prices*, Yale University Press.

Georgescu-Roegen N. (1971), *The Entropy Law and the Economic Process*, Harvard University Press.

Giannerini S., Goracci G. (2023), Entropy-based tests for complex dependence in economic and financial time series with the R package tseriesEntropy, *Mathematics*, 11(3), 757, DOI: 10.3390/math11030757.

Gervais S., Kaniel R., Mingelgrin D.H. (2001), The high-volume return premium, *The Journal of Finance*, 56(3), 877–919.

Granger C. (1980), Tests for causation – a personal viewpoint, *Journal of Economic Dynamics and Control*, 2, 329–352.

Guo W.Y. (2001), Maximum entropy in option pricing: a convex-spline smoothing method, *Journal of Futures Markets*, 21(9), 819–832.

Gulko L. (1997), Dartboards and asset prices: Introducing the entropy pricing theory, *Advances in Economics*, 12, 237–276.

Gulko L. (1999), The entropy theory of stock option pricing, *International Journal of Theoretical and Applied Finance*, 2(3), 331–355.

Gulko L. (2002), The entropy theory of bond option pricing, *International Journal of Theoretical and Applied Finance*, 5(3), 355–383.

Hilbert M., Darmon D. (2020), How complexity and uncertainty grew with algorithmic trading, *Entropy*, 22(5), 499, DO: 10.3390/e22050499.

Hou K., Xiong W., Peng L. (2009), *A tale of two anomalies: the implications of investor attention for price and earnings momentum*, unpublished working paper, Ohio State University, City University of New York, Princeton University.

Huang J., Yong W.A., Hong L. (2016), Generalization of the Kullback-Leibler divergence in the Tsallis statistics, *Journal of Mathematical Analysis and Applications*, 436(1), 501–512.

Jakimowicz A. (2020), The role of entropy in the development of economics, *Entropy*, 22(4), 452, DOI: 10.3390/e22040452.

Jones R.D., Redsun S.G., Frye R.E., Myers K.D. (2003), *The Maxwell Demon and market efficiency*, arXiv:physics/0311074.

Keskin Z., Aste T. (2020), Information-theoretic measures for nonlinear causality detection: application to social media sentiment and cryptocurrency prices, *Royal Society Open Science*, 7(9), 200863, DOI: 10.1098/rsos.200863.

Kim M., Sayama H. (2017), Predicting stock market movements using network science: an information theoretic approach, *Applied Network Science*, 2(35), 1–14, DOI: 10.1007/s41109-017-0055-y.

Kim S., Ku S., Chang W., Song J.W. (2020), Predicting the direction of US stock prices using effective transfer entropy and machine learning techniques, *IEEE Access*, 8, 111660–111682.

King M., Kay J. (2020), *Radical Uncertainty: Decision-Making for an Unknowable Future*, Hachette UK.

Kopczewska K. (2014), L-moments skewness and kurtosis as measures of regional convergence and cohesion, *Statistica Neerlandica*, 68(4), 251–266.

Kopczewska K., Churski P., Ochojski A., Polko A. (2017), *Measuring Regional Specialisation: A New Approach*, Springer.

Krishnan H., Nelken L. (2001), Estimating implied correlations for currency basket options using the maximum entropy method, *Derivatives Use, Trading & Regulation*, 7(1), 1–7.

Kullback S., Leibler R.A. (1951), On information and sufficiency, *The Annals of Mathematical Statistics*, 22(1), 79–86.

Liao J., Peng C., Zhu N. (2022), Extrapolative bubbles and trading volume, *The Review of Financial Studies*, 35(4), 1682–1722.

Maasoumi E., Racine J. (2002), Entropy and predictability of stock market returns, *Journal of Econometrics*, 107(1–2), 291–312.

Madiman M. (2008), On the entropy of sums, *2008 IEEE Information Theory Workshop*, IEEE.

Mandelbrot B. (1997), *Fractals, Hasard et Finance*, Flammarion.

Marschinski R., Kantz H. (2002), Analysing the information flow between financial time series, *The European Physical Journal B*, 30(2), 275–281.

Merton R.C. (1973), Theory of rational option pricing, *The Bell Journal of Economics and Management Science*, 4(1), 141–183.

Merton R.C. (1987), A simple model of capital market equilibrium with incomplete information, *The Journal of Finance*, 42(3), 483–510.

Natal J., Ávila I., Tsukahara V.B., Pinheiro M., Maciel C.D. (2021), Entropy: from thermodynamics to information processing, *Entropy*, 23(10), 1340, DOI: 10.3390/e23101340.

Neto D. (2022), Examining interconnectedness between media attention and cryptocurrency markets: a transfer entropy story, *Economics Letters*, 214, 110460, DOI: 10.1016/j.econlet.2022.110460.

Nichols J.M., Bucholtz F., Michalowicz J.V. (2013), Linearized transfer entropy for continuous second--order systems, *Entropy*, 15(8), 3186–3204.

Patra S., Hiremath G.S. (2022), An entropy approach to measure the dynamic stock market efficiency, *Journal of Quantitative Economics*, 20(2), 337–377.

Philippatos G.C., Wilson C.J. (1972), Entropy, market risk, and the selection of efficient portfolios, *Applied Economics*, 4, 209–220.

Riek R. (2020), Entropy derived from causality, *Entropy*, 22(6), 647, DOI: 10.3390/e22060647.

Rothenstein R. (2018), *Quantification of market efficiency based on informational-entropy*, arXiv:1812.02371.

Schinckus C. (2009), Economic uncertainty and econophysics, *Physica A: Statistical Mechanics and its Applications*, 388(20), 4415–4423.

Schreiber T. (2000), Measuring information transfer, *Physical Review Letters*, 85(2), 461.

Shannon C.E. (1948), A mathematical theory of communication, *The Bell System Technical Journal*, 27(3), 379–423.

Simon H.A. (1971), Designing organizations for an information-rich world, in: M. Greenberger (ed.), *Computers, Communications, and the Public Interest*, Johns Hopkins University Press.

Smith J. (2017), *Maximum entropy and information theory approaches to economics*, SSRN 3094757.

Syczewska E., Struzik Z. (2015), Granger causality and transfer entropy for financial returns, *Acta Physica Polonica A*, 127(3-A), A-129–A-135.

Tsallis C. (1988), Possible generalization of Boltzmann-Gibbs statistics, *Journal of Statistical Physics*, 52(1–2), 479–487, DOI: 10.1007/BF01016429.

Tsallis C., Levy S.V., Souza A.M., Maynard R. (1995), Statistical-mechanical foundation of the ubiquity of Lévy distributions in nature, *Physical Review Letters*, 75(20), 3589.

Wang Z. (2021), The high volume return premium and economic fundamentals, *Journal of Financial Economics*, 140(1), 325–345.

Wolfram S. (2023), *The Second Law: Resolving the Mystery of the Second Law of Thermodynamics*, Wolfram Media.

Yao C.Z., Li H.Y. (2020), Effective transfer entropy approach to information flow among EPU, investor sentiment, and stock market, *Frontiers in Physics*, 8, 206, DOI: 10.3389/fphy.2020.00206.

Zhou R., Cai R., Tong G. (2013), Applications of entropy in finance: a review, *Entropy*, 15(11), 4909–4931, DOI: 10.3390/e15114909.

Appendix

Table 1

Average returns and volatility (standard deviation) of periodical returns

Index	Time lag	2019	2020	2021	2022	2019	2020	2021	2022
		Average of returns				Standard deviation of returns			
WIG20	1 minute	-1.0e-06	0	1.0e-06	-1.0e-06	0.00017	0.00033	0.00022	0.00028
WIG20	1 hour	-4.2e-05	-8.0e-06	8.0e-06	-1.7e-05	0.00138	0.00274	0.00185	0.00240
WIG20	1 day	-9.7e-04	-1.7e-04	1.3e-04	-3.8e-04	0.00782	0.01314	0.01079	0.01309
S&P 500	1 minute	-1.1e-05	-8.0e-06	-3.0e-06	-6.0e-06	0.00018	0.00043	0.00022	0.00032
S&P 500	1 hour	6.0e-06	5.0e-05	9.0e-06	-1.8e-05	0.00140	0.00335	0.00163	0.00244
S&P 500	1 day	4.5e-04	1.1e-03	3.0e-04	-2.9e-04	0.00715	0.01671	0.00725	0.01184

Table 2

Fat-tails of distributions – skewness and kurtosis of returns

Index	Time lag	Skewness of returns				Kurtosis of returns			
		2019	2020	2021	2022	2019	2020	2021	2022
WIG20	1 minute	-0.48	0.36	1.50	0.06	79.4	192.7	138.6	70.2
WIG20	1 hour	-0.72	0.67	0.82	0.17	32.5	52.8	40.5	21.9
WIG20	1 day	-0.41	-0.16	1.26	0.02	6.35	7.37	15.59	4.49
S&P 500	1 minute	0.23	0.65	-0.10	-1.89	94.8	94.8	33.8	16.0
S&P 500	1 hour	-0.75	0.75	0.09	-0.20	29.08	46.8	21.8	22.3
S&P 500	1 day	-1.48	0.19	-0.21	0.02	11.06	15.4	4.6	5.5

Table 3

Market liquidity – the average transaction volume

Index	Time lag	Average volume			
		2019	2020	2021	2022
WIG20	1 minute	401	338	224	512
WIG20	1 hour	24 084	20 308	12 369	32 276
WIG20	1 day	576 453	486 109	307 313	774 260
S&P 500	1 minute	28 876	100 646	75 269	92 262
S&P 500	1 hour	1 732 566	6 037 990	5 225 967	5 491 875
S&P 500	1 day	41 467 973	144 529 194	127 966 114	131 383 871

Table 4

Market memory and asymmetry – fit of q-normal (Tsallis) distribution to returns

Index	Time lag Δt	2019				2020			
		q^-	q	q^+	$\lvert\delta q\rvert/q$	q^-	q	q^+	$\lvert\delta q\rvert/q$
WIG20	1 minute	1.5321	1.5170	1.4702	4.1%	1.4887	1.5025	1.5135	1.7%
WIG20	1 hour	1.3864	1.3931	1.3938	0.5%	1.4569	1.4504	1.4404	1.1%
WIG20	1 day	1.3333	1.3683	1.3847	3.8%	1.4499	1.3936	1.3129	9.8%
S&P 500	1 minute	1.5555	1.5663	1.5675	0.8%	1.5613	1.5661	1.5791	1.1%
S&P 500	1 hour	1.5475	1.5388	1.4853	4.0%	1.4889	1.5132	1.5421	3.5%
S&P 500	1 day	1.3672	1.4050	1.4495	5.9%	1.3837	1.4877	1.5398	10.5%
Index	**Time lag Δt**	**2021**				**2022**			
		q^-	q	q^+	$\lvert\delta q\rvert/q$	q^-	q	q^+	$\lvert\delta q\rvert/q$
WIG20	1 minute	1.4683	1.4878	1.4996	2.1%	1.3478	1.3739	1.4119	4.7%
WIG20	1 hour	1.5028	1.4856	1.4227	5.4%	1.0000	1.3690	1.3910	28.6%
WIG20	1 day	1.4735	1.4134	1.3402	9.4%	1.3766	1.2457	1.0000	30.2%
S&P 500	1 minute	1.4796	1.4889	1.4416	2.6%	1.3828	1.3864	1.4011	1.3%
S&P 500	1 hour	1.5002	1.5035	1.5076	0.5%	1.4786	1.4882	1.5006	1.5%
S&P 500	1 day	1.3185	1.3137	1.3522	2.6%	1.2795	1.1972	1.0000	23.3%

Table 5

Interpretation of market disorder based on one-dimensional entropy measures

Disorder of the market		Volume	
		low entropy	high entropy
Returns	low entropy	Returns: inefficient market, highly predictable with low risk, limited dispersion of returns Volume: illiquid (shallow) market	Returns: inefficient market, highly predictable with low risk, limited dispersion of returns Volume: liquid (deep) market
Returns	high entropy	Returns: efficient market, difficult to predict, predictions with high risk, high dispersion of returns (toward uniform distribution) Volume: illiquid (shallow) market	Returns: efficient market, difficult to predict, predictions with high risk, high dispersion of returns (toward uniform distribution) Volume: liquid (deep) market

Table 6
Market disorder – Shannon entropy of returns and transactions volume

Index	Time lag	Shannon entropy of returns				Shannon entropy of volume			
		2019	2020	2021	2022	2019	2020	2021	2022
WIG20	1 minute	0.0002	0.0031	0.0007	0.0008	0.6786	0.6695	0.5607	0.5721
WIG20	1 hour	0.0112	0.0961	0.0320	0.0784	0.7652	0.8011	0.7484	0.7623
WIG20	1 day	0.5759	0.9928	0.7642	1.1083	1.5023	1.2561	1.0757	1.3423
S&P 500	1 minute	0.0003	0.0054	0.0001	0.0010	0.7849	1.3137	1.1732	1.2108
S&P 500	1 hour	0.0144	0.1173	0.0233	0.0547	1.2060	0.7640	0.8460	0.9737
S&P 500	1 day	0.1420	0.4883	0.1077	0.3920	0.3775	1.4033	1.2976	1.1603

Table 7
Kullback-Leibler Divergence for a given year calculated between returns and liquidity (volume)

Index	Time lag	KLD for a given year between return and liquidity			
		2019	2020	2021	2022
WIG20	1 minute	1.273	6.602	1.386	7.313
WIG20	1 hour	0.162	2.709	0.166	0.166
WIG20	1 day	1.215	1.810	2.372	1.355
S&P 500	1 minute	1.221	4.907	0.693	0.655
S&P 500	1 hour	0.140	4.585	1.351	1.123
S&P 500	1 day	0.219	1.479	0.432	0.536

Table 8
KLD matrices between years for return and liquidity jointly

KLD		WIG20				S&P 500			
		2019	2020	2021	2022	2019	2020	2021	2022
1 minute	2019	0.00000	0.00444	0.03096	0.04393	0.00000	0.00793	0.00196	0.00334
	2020	**0.00444**	0.00000	0.04315	0.05846	**0.00793**	0.00000	0.00217	0.00112
	2021	0.03096	**0.04315**	0.00000	0.00114	0.00196	**0.00217**	0.00000	0.00019
	2022	0.04393	0.05846	**0.00114**	0.00000	0.00334	0.00112	**0.00019**	0.00000
1 hour	2019	0.00000	0.00285	0.00601	0.00014	0.00000	0.00539	0.00433	0.00221
	2020	**0.00285**	0.00000	0.00064	0.00404	**0.00539**	0.00000	0.00006	0.00071
	2021	0.00601	**0.00064**	0.00000	0.00777	0.00433	**0.00006**	0.00000	0.00037
	2022	0.00014	0.00404	**0.00777**	0.00000	0.00221	0.00071	**0.00037**	0.00000
1 day	2019	0.00000	0.00215	0.00481	0.00214	0.00000	0.01765	0.01215	0.01074
	2020	**0.00215**	0.00000	0.00176	0.00127	**0.01765**	0.00000	0.00063	0.00190
	2021	0.00481	**0.00176**	0.00000	0.00496	0.01215	**0.00063**	0.00000	0.00077
	2022	0.00214	0.00127	**0.00496**	0.00000	0.01074	0.00190	**0.00077**	0.00000

Table 9
Kullback-Leibler Divergence in the Tsallis statistics – comparison of returns between years

TKLD		WIG20				S&P 500			
		2019	2020	2021	2022	2019	2020	2021	2022
1 minute	2019	0.0000	0.0002	0.0007	0.0102	0.0000	0.0000	0.0067	0.0234
	2020	**0.0002**	0.0000	0.0002	0.0076	**0.0000**	0.0000	0.0066	0.0233
	2021	0.0007	**0.0002**	0.0000	0.0056	0.0067	**0.0066**	0.0000	0.0047
	2022	0.0102	0.0076	**0.0056**	0.0000	0.0234	0.0233	**0.0047**	0.0000
1 hour	2019	0.0000	0.0013	0.0039	0.0002	0.0000	0.0007	0.0012	0.0023
	2020	**0.0013**	0.0000	0.0007	0.0024	**0.0007**	0.0000	0.0001	0.0005
	2021	0.0039	**0.0007**	0.0000	0.0057	0.0012	**0.0001**	0.0000	0.0002
	2022	0.0002	0.0024	**0.0057**	0.0000	0.0023	0.0005	**0.0002**	0.0000
1 day	2019	0.0000	0.0002	0.0006	0.0034	0.0000	0.0032	0.0023	0.0100
	2020	**0.0002**	0.0000	0.0001	0.0052	**0.0032**	0.0000	0.0112	0.0255
	2021	0.0006	**0.0001**	0.0000	0.0071	0.0023	**0.0112**	0.0000	0.0027
	2022	0.0034	0.0052	**0.0071**	0.0000	0.0100	0.0255	**0.0027**	0.0000

Eksploracja dynamiki rynku akcji: podejście oparte na dwuwymiarowej entropii w przestrzeni zwroty/wolumen obrotów

Streszczenie

Badanie rozpoczyna się od eksperymentu myślowego: czy na podstawie próbek finansowych szeregów czasowych, bez znajomości okresów, z których pochodzą, można zidentyfikować te z okresu pandemii oraz te sprzed pandemii, bazując na ich unikalnych cechach lub statystykach? Okazało się, że trudno jest znaleźć statystykę, która jednoznacznie pozwalałaby przypisać próbki do jednego z tych dwóch okresów.

Zachowanie rynków finansowych w trakcie pandemii było opisywane jako dziwne, nieprzewidywalne i nieadekwatne w odniesieniu do realnej gospodarki, co sugeruje, że entropia, jako miara nieuporządkowania, mogłaby być rozsądnym kryterium odróżnienia tych okresów. Entropia była już wcześniej stosowana do badania niepewności na rynkach finansowych. W prezentowanym badaniu rynek finansowy jest traktowany inaczej – jako system informacyjny, który generuje dwa równoczesne sygnały dla inwestorów: zwroty i wolumen obrotów, dlatego miara entropii powinna uwzględniać oba te aspekty. Dotychczasowe badania koncentrowały się na relacji między cenami/zwrotami a wolumenem obrotów, głównie w kontekście prognozowania zmian cen na podstawie wolumenu. Niniejsze badanie wyróżnia się, proponując zastosowanie miar entropii łącznej do analizy wpływu pandemii na statystyki rynków finansowych oraz sprawdzenie, czy możliwe jest przypisanie charakterystycznych cech określonym okresom.

Główne pytanie badawcze brzmi: czy miary entropii mogą służyć jako narzędzie do rozróżniania okresów przed pandemią i w trakcie pandemii na rynkach finansowych, biorąc pod uwagę dwie kluczowe zmienne charakteryzujące aktywa finansowe: zwroty i wolumen obrotów?

Badanie ma charakter eksploracyjny, co oznacza, że jego celem jest odkrywanie i zrozumienie wzorców oraz zależności w danych, które do tej pory nie zostały dobrze zdefiniowane w literaturze. Podejście to umożliwia elastyczne traktowanie problemu badawczego i formułowanie nowych hipotez, które mogą być testowane w przyszłych pracach. Analiza miar entropii pozwala odkryć wzajemne relacje między zwrotem a wolumenem obrotów, zwłaszcza w kontekście wstrząsów makroekonomicznych. Oczekiwaliśmy, że entropia i jej składowe będą znacząco wyższe w okresach szoków, a w okresach spokojniejszych ulegną obniżeniu, oraz że wzajemne relacje między poszczególnymi składnikami łącznej entropii będą różne w analizowanych okresach. Badanie zakłada, że rynki finansowe nie są systemami izolowanymi oraz że zwiększony napływ zewnętrznych informacji wpływa na ceny i wolumeny obrotów, wprowadzając większe nieuporządkowanie w ich kształtowaniu się.

Analiza entropii dotyczy dwóch indeksów giełdowych: amerykańskiego S&P 500 oraz polskiego WIG20, w czterech wybranych latach reprezentujących okresy: przed pandemią, w trakcie pandemii, postpandemiczny, a także w czasie wojny w Ukrainie. Do badania wykorzystano dane o trzech różnych częstotliwościach: minutowej, godzinowej i dziennej. Wykorzystano metody oparte na entropii, takie jak rozkład Tsallisa i Dywergencję Kullbacka-Leiblera, do analizy jedno- i dwuwymiarowej.

Wstępne analizy wskazują na różnice w miarach entropii między analizowanymi okresami, co sugeruje, że entropia może być użytecznym wskaźnikiem oceny wpływu zewnętrznych szoków

na rynki finansowe. Wyniki sugerują, że entropia i jej komponenty mogą stanowić istotne narzędzia analizy dynamiki rynków finansowych, zwłaszcza w okresach niestabilności gospodarczej. Zmiany w zakresie entropii zwrotów i wolumenu obrotów mogą być także wartościową zmienną w algorytmach prognozowania uczenia maszynowego, odzwierciedlającą pełną nieprzewidywalność systemu.

Należy jednak uwzględnić pewne ograniczenia wynikające z zastosowanych metod estymacji entropii oraz zakresu dostępnych danych, co może wpłynąć na interpretację wyników. Badanie ma charakter eksploracyjny i jego wyniki wymagają dalszej weryfikacji w szerszym kontekście rynków finansowych. Problemem z miarami entropii jest trudność ich wykorzystania w standardowych modelach finansowych. Przede wszystkim uzyskane wyniki są trudne do jednoznacznej i łatwej interpretacji, co stanowi wyzwanie w modelach finansowych, szczególnie w tych tworzonych na potrzeby instytucji regulacyjnych. Entropia nie zastąpi w modelach finansowych zmienności wyrażonej przez odchylenie standardowe, choć badania empiryczne sugerują, że może być lepszą miarą zmienności w analizach portfelowych. Wynika to z niechęci do odchodzenia od łatwo interpretowalnych miar statystycznych.

Sytuacja staje się jeszcze bardziej skomplikowana w przypadku miar entropii wzajemnej, które mogą uchwycić częściowe i nieliniowe zależności między analizowanymi zmiennymi. W przypadku danego rodzaju aktywów finansowych relacja zwrot–wolumen może mieć określony kierunek dopiero po przekroczeniu pewnego progu wartości zmiennych. Relacja ta może być również asymetryczna, z różnymi siłami oddziałującymi w okresach wzrostów i spadków. Na przykład wyprowadzenie miar entropii z rozkładu Tsallisa pozwala na oszacowanie asymetrii zwrotów, gdzie zmienność mierzona entropią reaguje inaczej na zwroty dodatnie i ujemne. Entropia wzajemna i jej składowe mogą uchwycić asymetrię w relacjach między zwrotami a wolumenem obrotu, a także asymetrię idiosynkratyczną, co oznacza, że relacje te mogą wynikać nie tylko z czynników rynkowych, ale także ze specyficznych zdarzeń dotyczących danej firmy lub jej cech.

Miary entropii dostarczają informacji innych niż np. indeks VIX. Są bardziej złożone, ale z uwagi na ich warunkową i idiosynkratyczną naturę trudno jest jednoznacznie zinterpretować ich reakcję na wstrząsy zewnętrzne oraz ustalić stałą relację pomiędzy nimi a zachowaniem rynków. Miary te będą najbardziej użyteczne w modelach prognostycznych lub portfelowych, działających jak algorytmiczne „czarne skrzynki" (LM, AI), co czyni je interesującymi zmiennymi dla handlu algorytmicznego, ale niekoniecznie dla tradycyjnych modeli finansowych.

Słowa kluczowe: rynki finansowe, zwrot i wolumen obrotu, entropia, Dywergencja Kullbacka-Leiblera, rozkład Tsallisa