Quantifying horizon dependence of asset prices: a cluster entropy approach

Linda Ponta\textsuperscript{1}, Anna Carbone\textsuperscript{2},

\textsuperscript{1}LIUC-Università Cattaneo, corso Giacomo Matteotti 22, 21053 Castellanza, Italy
\textsuperscript{2}Politecnico di Torino, corso Duca degli Abruzzi 24, 10129 Torino, Italy

Abstract

Market dynamic is studied by quantifying the dependence of the entropy $S(\tau, n)$ of the clusters formed by the series of the prices $p_t$ and its moving average $\tilde{p}_{t,n}$ on temporal horizon $M$. We report results of the analysis performed on high-frequency data of the Nasdaq Composite, Dow Jones Industrial Avg and Standard & Poor 500 indexes downloaded from the Bloomberg terminal www.bloomberg.com/professional. Both raw and sampled data series have been analysed for a broad range of horizons $M$, varying from one to twelve months over the year 2018. A systematic dependence of the cluster entropy function $S(\tau, n)$ on the horizon $M$ has been evidenced in the analysed assets. Hence, the cluster entropy function is integrated over the cluster $\tau$ to yield a synthetic indicator of price evolution: the Market Dynamic Index $I(M, n)$. Moreover, the Market Horizon Dependence defined as $H(M, n) = I(M, n) - I(1, n)$ is calculated and compared with the values of the horizon dependence of the pricing kernel with different representative agent models obtained by a Kullback-Leibler entropy approach.

1 Introduction

Entropy, as a tool to quantify heterogeneity and evolution in complex systems, has found a number of applications in different contexts\cite{1-6}.

In economics and finance, entropy concepts have been exploited for portfolio selection to complement and outperform traditional methods based on Markowitz covariance and Sharpe single-index models\cite{7-20}. An information theoretical tool has been recently developed yielding the weights of the efficient portfolio\cite{21,22} by using the procedure based on the cluster entropy estimated via the detrending moving average algorithm proposed in\cite{23,24}. Interestingly, the cluster entropy of the volatility series takes values which sensibly depend on the market, as opposed to the entropy of the prices, which was shown to be approximately invariant across the markets. The Market Heterogeneity Index, defined as the integral of the cluster entropy functional of the volatility, allowed a direct comparison with the portfolio weights obtained by the Sharpe ratio approach. The cluster entropy approach has the advantage of not requiring a specific distribution of returns, such as a symmetric Gaussian distribution, which is quite elusive in real-world financial assets, thus hindering, in principle, the application of Markowitz-based portfolio models.

Moreover, there is an increasing interest to implement entropy-derived tools and concepts to assess the validity and shed light on fundamental aspects of asset pricing models beyond cross-market portfolio optimization\cite{26,30}. Economy-wide shocks,
which are intrinsically not diversifiable, are gaining increasing importance due to the growing connectedness of the assets. The propagation of these shocks cannot be averaged out by simply diversifying investments, thus even the best selection of portfolio assets might fail in keeping investors safe. The development of accurate instruments accounting for macroeconomic ingredients to risk is of key interest for quantitative finance. Asset pricing models provide risk estimates by quantifying market evolution in terms of a stochastic function: the pricing kernel \( m_t \). Equilibrium prices \( p_t \) of traded securities can be represented as the conditional expectation of the discounted future payoff \( z_t \):

\[
p_t = E \left[ \frac{m_{t+1}}{m_t} z_{t+1} \right],
\]

(1)

and \( m_{t+1}/m_t \) is known as the stochastic discount factor. The pricing kernel \( m_t \) is factorizable into a function of the consumption growth \( \mu_{t+1} \) times a model specific term \( \psi_t \):

\[
m_t = \mu_{t+1} \psi_t.
\]

(2)

The standard consumption-based asset pricing model identifies the pricing kernel as a simple parametric function of the consumption growth \( C_t \). In this framework, with time-separable power utility representative agent models, the function \( \mu_{t+1} \) is simply proportional to \( \Delta C_t = \log(C_t/C_{t-1}) \). More sophisticated agent behaviours have been suggested to explain puzzling phenomena such as amplitude and cross-sectional dispersion of returns among different categories of financial assets, equity premia and risk-free rates.

Dispersion and dynamic of the pricing kernel with different representative agent behaviour have been modelled by using the Kullback-Leibler entropy [29]. The work builds on and extends the results reported in the seminal paper by Hansen and Jagannathan mainly addressed to quantify standard deviation and volatility and, ultimately, define the bounds of pricing kernels. A lower bound was provided for the volatility of the permanent component of asset pricing kernels, which showed that stochastic discount factors need to be very volatile to be consistent with high Sharpe ratios [26]. A relative entropy minimization based on the Kullback-Leibler divergence approach to extract the model dependent term \( \mu_{t+1} \) has been put forward in [30] with the ultimate purpose to quantify the minimum amount of extra information to be added to the standard pricing kernel models for reproducing asset return correctly. The Kullback-Leibler divergence between the probability distribution functions of the components \( \mu_{t+1} \) and \( \psi_t \) has been used as criterion to estimate the deviation of \( m_{t+1} \) with respect to the simple consumption flow growth model. It is argued that the Kullback-Leibler divergence criterion is equivalent to maximize the entropy of the fundamental component of the pricing kernel [30].

Motivated by the growing interest in modelling price dynamics and understanding fundamental sources of risk, in this work we exploit the cluster entropy approach to the purpose of understanding the intrinsic dynamic of asset prices. As mentioned in the abstract, the present work builds upon and extends the study [22] that was essentially devoted to extract the portfolio weights from the cluster entropy of the volatility which exhibits sensible cross-market variations. Conversely, the cluster entropy of the prices has been found to be almost invariant across the markets, suggesting that it could be more deeply connected to the macroeconomic fundamental properties rather than cross market variations.

Hence, in this work, we focus on price evolution by performing the cluster entropy analysis on several assets over multiple horizons. This aspect was not included in our previous study [22] that reported results obtained over a constant time interval.
A systematic dependence of the cluster entropy of the asset prices over time has been observed. For the sake of clarity, the main relationships relevant to the cluster entropy approach adopted in this paper will be shortly recalled in Section 2. Then the data set of financial assets, whose results are reported in this manuscript, is described in Section 3. In Section 4 the results of the entropy of the prices series as a function of the horizon $M$ are reported, together with a comparison with the results obtained by simulating the pricing kernel with different representative agent models.

2 Methods

The ability of the cluster entropy approach to quantify the intrinsic dynamic of the prices is proved by analysing several assets. For the sake of simplicity in this work we report the results obtained on the three markets described in Table 1. In this section, before discussing in details the computational results, we briefly recall the main definitions and equations used.

The cluster entropy is obtained by taking the intersection of the asset prices $p_t$ and its moving average $\tilde{p}_{t,n}$ for different moving average window $n$. For each window $n$, the subsets $\{p_t : t = s, ..., s - n\}$ between two consecutive intersections are considered. The subsets are named clusters. The clusters are exactly defined as the portions of the series between death/golden crosses according to the technical trading rules. Therefore, the information content has a straightforward connection with the trader’s perspective on the price and volatility series. Then, the clusters are ranked according to their characteristic size, the duration $\tau$. The probability distribution function $P(\tau, n)$ of the cluster duration is obtained. The present approach directly yields either power-law or exponential distributed cluster distributions, thus enabling us to separate the sets of inherently correlated/uncorrelated blocks along the sequence.

The continuously compounded return is defined by:

$$r_t = p_t - p_{t-h} \ ,$$

where $p_t$ is the price at the time $t$, with $0 < h < t < N$ and $N$ the maximum length of the time series. Alternatively, one can consider the log-return defined as:

$$r_t = \log p_t - \log p_{t-h} \ .$$

The approach adopted in this work builds upon the idea of Claude Shannon to quantify the expected information contained in a message extracted from a sequence $\{x_t\}$ by using the entropy functional:

$$S[P(x_t)] = \int_X p(x_t) \log p(x_t) dx_t \ ,$$

with $P$ a probability distribution function associated with the sequence $\{x_t\}$. For discrete sets, Eq. (5) reduces to:

$$S[P(x_t)] = \sum_X p(x_t) \log p(x_t) \ .$$

Consider the time series $\{x_t\}$ of length $N$ and the moving average $\{\tilde{x}_{t,n}\}$ of length $N - n$ with $n$ the moving average window. The function $\{\tilde{x}_{t,n}\}$ generates, for each $n$, a partition $\{C\}$ of non-overlapping clusters between two consecutive intersections of $\{x_t\}$ and $\{\tilde{x}_{t,n}\}$. Each cluster $j$ has duration:

$$\tau_j \equiv \|t_j - t_{j-1}\|$$
where the instances \( t_{j-1} \) and \( t_j \) refer to two subsequent intersections. The probability distribution function \( P(\tau, n) \) can be obtained by ranking the number of clusters \( N(\tau_1, n), N(\tau_2, n), \ldots, N(\tau_j, n) \) according to their length \( \tau_1, \tau_2, \ldots, \tau_j \) for each \( n \). A stationary sequence of clusters \( C \) is generated with probability distribution function varying as \( P(\tau, n) \sim \tau^{-\alpha}F(\tau, n) \),

\[
P(\tau, n) \sim \tau^{-\alpha}F(\tau, n)
\]

that by using Eq. (8) simplifies to:

\[
S(\tau, n) = S_0 + \log \tau + \frac{\tau}{n}
\]

where \( S_0 \) is a constant, \( \log \tau \) and \( \tau/n \) are related respectively to the terms \( \tau^{-\alpha} \) and \( F(\tau, n) \). The minimum value of the entropy is obtained for the fully ordered (deterministic) set of clusters with duration \( \tau = 1 \). Eq. (10) in the limit \( n \sim \tau \to 1 \) and \( S_0 \to -1 \) reduces to \( S(\tau, n) \to 0 \). Conversely, the maximum value of the entropy \( S(\tau, n) = \log N^\alpha \) is obtained with \( n \sim \tau \to N, N \) being the maximum length of the sequence. This condition corresponds to the maximum randomness (minimum information) carried by the sequence, when a single longest cluster is obtained coinciding with the whole series. For a fractional Brownian motion, the exponent \( \alpha \) is equal to the fractal dimension \( D = 2 - H \) with \( H \) the Hurst exponent of the time series. The term \( \log \tau \) can be thus interpreted as a generalized form of the Boltzmann entropy \( S = \log \Omega \), where \( \Omega = \tau^D \) corresponds to the fractional volume occupied by the fractional random walker. The term \( \tau/n \) represents an excess entropy (excess noise) added to the intrinsic entropy term \( \log \tau^D \) by the partition process. It depends on \( n \) and is related to the finite size effect discussed above.

We stress the difference between the time series partitions obtained either by using equal size boxes or moving average clusters. For equal size boxes, the excess noise term \( \tau/n \) vanishes (as it becomes a constant that can be included in the constant term) thus the entropy reduces to the logarithmic term as found in Ref. [3], which corresponds to the intrinsic entropy of an ideal fractional random walk. When a moving average partition is used, an excess entropy term \( \tau/n \) emerges accounting for the additional heterogeneity introduced by the random partitioning process operated by the moving average intersections.

To univocally quantify market properties through the entropy Eq. (10), a cumulative information measure has been defined as follows:

\[
I(n) = \int_0^{\tau_{\text{max}}} S(\tau, n) d\tau
\]

which, for discrete sets, reduces to:

\[
I(n) = \sum_{\tau} S(\tau, n)
\]

The function \( I(n) \) has been used to quantify cross-market heterogeneity in [22]. The cluster entropy of the volatility \( v_T \) was integrated over the cluster duration \( \tau \) to the purpose of obtaining the weights of the optimal portfolio.

In this work, the function \( I(n) \) will be used to quantify the intrinsic market dynamic. The cluster entropy of the prices will be integrated over the cluster duration \( \tau \) to the purpose of obtaining the horizon dependence.
3 Data

Prices of market indices traded in the US, namely NASDAQ, DJIA and S&P500 are investigated. Data sets have been downloaded from the terminal www.bloomberg.com/professional. For each index, the data set includes tick-by-tick prices \( p_t \) from January to December 2018. The main characteristics of the three asset are summarized in Table 1. The last column in the table reports the length of each index referred to the year 2018. Different temporal horizons have been considered up to twelve integer monthly multiples of one-month period. The lengths of the series referred to the twelve periods are reported for each index in Table 2.

To the purpose of performing a set of analysis at constant lengths, the raw data series have been sampled to obtain data series with equal length. The sampling frequency is defined for each series by dividing the length of the series corresponding to the longest horizon by the minimum, rounding the ratio to the nearest whole, that is used to sample the raw data. Consider for example the S&P500 market (3rd column in Table 2). The minimum value of the length is that at \( M = 1 \) (January with \( N = 516635 \)) and the maximum value is longest horizon of interest (for example \( N = 5180006 \) for horizon \( M = 10 \) equal to ten months from January to October). As the sampling frequency is different for each series we consider the minimum value to perform the analysis with the same length.

Furthermore for the sake of training the algorithm, we provide a set of figures showing the results obtained with artificially generated series of different lengths. The different lengths are the same as those of the financial markets. For example for the curves in Figs. S1 Figure (Supplementary Material) the values of the different lengths have been taken equal to those of the NASDAQ (first column of Table 2). The artificial series can be generated by means of the FRACLAB tool available at: https://project.inria.fr/fraclab/.

4 Results

Probability distribution \( P(\tau, n) \) and entropy \( S(\tau, n) \) functions have been calculated for a large set of prices series by means of the procedure summarized in Section 2. The series of the NASDAQ, DJIA and S&P500 indexes described in Section 3 have been used for the investigation.

Figs. 1 show the cluster entropy \( S(\tau, n) \) calculated by using raw data prices. In particular, the plots refer to one month of data (\( M = 1 \)). The series lengths are \( N = 586866 \), \( N = 516644 \) and \( N = 516635 \) respectively for NASDAQ, DJIA and S&P500 as given in Table 2.

Figs. 2 show the cluster entropy \( S(\tau, n) \) calculated by using raw data prices, as in Figs. 1 but here the series refer to the horizon of twelve months (\( M = 12 \)). The series lengths are \( N = 6982017 \), \( N = 5749145 \) and \( N = 6142443 \) respectively for NASDAQ, DJIA and S&P500 as one can check in the last row of Table 2.

Figs. 3 show the cluster entropy \( S(\tau, n) \) calculated by using the prices series of the sampled data. The plots refer to first month of data (\( M = 1 \)). The series have same length \( N = 492035 \).

Figs. 4 show the cluster entropy \( S(\tau, n) \) calculated by using the prices series of the sampled data. The plots refer to twelve months (\( M = 12 \)). The series have same length \( N = 492035 \).

Different curves in each figure correspond to moving average values varying from \( n = 30 \) s, \( n = 50 \) s, \( n = 100 \) s, \( n = 150 \) s, \( n = 200 \) s up to \( n = 1500 \) s with step 100s.

One can note that the entropy curves exhibit a behaviour consistent with Eq. (10). At small values of the cluster duration \( \tau \leq n \), entropy behaves as a logarithmic
function. At large values of the cluster duration $\tau \geq n$ the curves increase linearly with the term $\tau/n$ dominating. $S(\tau, n)$ is $n$-invariant for small values of $\tau$, while its slope decreases as $1/n$ at larger $\tau$, as expected according to Eq. (10), meaning that clusters with duration $\tau > n$ are not power-law correlated, due to the finite-size effects introduced by the partition with window $n$. Hence, they are characterized by a value of the entropy exceeding the curve $\log \tau^D$, which corresponds to power-law correlated clusters. It is worthy to remark that clusters with same duration $\tau$ can be generated by different values of the moving average window $n$. At a constant value of $\tau$, larger entropy values are obtained as $n$ increases.

The entropy $S(\tau, n)$ of the NASDAQ, DJIA and S&P500 prices (shown in Figs. 1, Figs. 2, Figs. 3 and Figs. 4) is representative of a quite general behaviour observed in several markets analysed by using the proposed cluster entropy approach.

In the following, we will discuss how to quantify the horizon dependence of the asset prices by using the cluster entropy function $S(\tau, n)$ estimated over different periods $M$. To this purpose, we use the cumulative information measure $I(n)$, i.e. the function defined in Eq. (11).

In this work, the quantity $I(M, n)$ is calculated by using the values of the entropy $S(\tau, n)$ of the asset prices $p_t$ over several periods (namely $M$ ranging from one to twelve months) by using both raw and sampled data. The first period ($M = 1$) of the price sequences is taken in correspondence of January 2018 for all the assets. Multiple period sequences have been built by considering $M = 2$ (January and February 2018) and, so on, up to $M = 12$ (whole year from January to December 2018). Details concerning lengths of the series are reported in Table 2.

The cumulative information measure $I(M, n)$ has been plotted in Fig. 5 for the prices of the NASDAQ, DJIA and S&P500. One can observe a dependence of the function $I(M, n)$ at different $M$ horizons. $I(M, n)$ is the same for all $M$ implying that the horizon dependence $H(M, n)$ is negligible at small scales (small $n$/small $\tau$ values). Conversely, at large $n$ values, i.e. with a broad range of cluster lengths $\tau$ spanning more than one decades of values in the power law distribution, a horizon dependence $H(M, n)$ varying with $M$ is found.

For identically distributed sequences of clusters, $I(M, n)$ does not change with $M$ regardless of the value of $n$. This, has been shown in the Figs. S1 (Supplementary Material) where the cluster entropy $S(\tau, n)$ of artificially generated series (fractional random walks) are shown. One can note that the curves are practically unchanged at varying horizons $M$ and cluster duration $\tau$. The departure from the iid case can be taken as a measure of price dynamics.

Furthermore, by comparing the figures corresponding to the different assets a dependence of the function $I(n)$ is observed. In the case of the NASDAQ the variation seems larger than for the S&P500, and even larger than for the DJIA.

Discussion

Next, the main results of the analysis of the cluster entropy $S(\tau, n)$ and the cumulative information measure $I(M, n)$ are compared with the results obtained by using information theoretical approaches by other authors.

To build a cluster entropy index of horizon dependence, i.e. a synthetic numerical parameter with the ability to provide an estimate of the horizon dependence, we consider the entropy integral $I(n)$ defined by Eq. (12) at one-period ($M = 1$) and at multiples of one period $M$ respectively defined as $I(1, n)$ and $I(M, n)$. The quantity $I(M, n)$, defined above on the basis of Eq. (12) is called Market Dynamic Index.

To the purpose of comparing our results with those of paper [29], the horizon
dependence $H(M, n)$ is calculated as:

$$H(M, n) = I(M, n) - I(1, n) \quad .$$

(13)

Values of Market Dynamic Index $I(M, n)$ and horizon dependence $H(M, n)$ calculated by using the NASDAQ, DJIA and S&P500 data are reported in Table 3. The quantity $I(1, n) = I(1)$ is a reference value of the one-period entropy (lower bound). It is taken as $I(1) = 0.0049$, $I(1) = 0.0214$ and $I(1) = 0.0197$ respectively for power utility, recursive utility and difference habit agent models of the consumption growth following [29]. The value $I(12, n)$ has been obtained from the curves in Figs. 5 for the prices of the NASDAQ, DJIA and S&P500. $H(12, n)$ is the difference between $I(12, n)$ and $I(1, n)$ on account of Eq. (13).

Next, the values of the horizon dependence obtained by using the cluster entropy will be checked against those obtained by using different representative agent models for the definition of the pricing kernel in [29]. The pricing kernel dynamics has been quantified by a measure of entropy dependence on the investment horizon for popular asset pricing models. The pricing kernel accounts for the stochastic dynamic evolution of asset returns, which in their turn contain information about the pricing kernel. The analysis is based on the Kulback-Leibner divergence (also known as relative entropy) of the true probability distribution of the prices with respect to the risk-adjusted probability. On account of those results, it was argued that a realistic asset pricing model should have substantial one-period entropy and modest horizon dependence to justify equity mean excess returns and bond yields at once.

The Kullback-Leibler (KL) divergence of the continuous probability measure $p(x)$ with respect to some probability measure $p^*(x_t)$, writes:

$$KL(P||P^*) = \int_X p(x_t) \log \left( \frac{p(x_t)}{p^*(x_t)} \right) dx_t$$

(14)

Eq. (14) can be interpreted as the expectation of the function $\log p(x_t)/p^*(x_t)$ with respect to the probability $p(x_t)$:

$$KL(P||P^*) = E \left[ \log \left( \frac{p(x_t)}{p^*(x_t)} \right) \right]$$

(15)

It can be easily shown that the relative entropy Eq. (14) reduces to Eq. (5) for constant probability $p^*(x_t)$.

Investigation of asset price dispersion and dynamics has been put forward by using a variant of the Kullback-Leibner (KL) divergence of the pricing kernels $m_{t,t+n}$ expressed in terms of the ratio between the true and risk-adjusted distribution [29]. Different representative agent models have been considered to quantify the horizon dependence $H(M)$ [29]:

$$H(M) = I(M) - I(1)$$

(16)

with the quantity $I(M)$ defined as:

$$I(M) = \frac{EL_t(m_{t,t+M})}{M} \quad .$$

(17)

where $EL_t(m_{t,t+M})$ is defined as the average of the relative entropy of the pricing kernel, and $I(1)$ is calculated at the month $M = 1$.

A summary of the horizon dependence obtained by estimating the KL entropy with pricing kernels generated by different representative agent models according to the approach of [29] is reported in Table 4.
References

[1] Crutchfield JP. Between order and chaos. Nature Physics. 2012;8(1):17–24.

[2] Bandt C, Pompe B. Permutation entropy: a natural complexity measure for time series. Phys Rev Lett. 2002;88(17):174102.

[3] Grassberger P, Procaccia I. Characterization of strange attractors. Phys Rev Lett. 1983;50(5):346.

[4] Marcon E, Scotti I, Hérault B, Rossi V, Lang G. Generalization of the partitioning of Shannon diversity. PLoS ONE. 2014;9(3):e90289.

[5] Karpierz M, Fronczak P, Fronczak A. International trade network: fractal properties and globalization puzzle. Phys Rev Lett. 2014;113(24):248701.

[6] Rubido N, Grebogi C, Baptista MS. Entropy-based generating Markov partitions for complex systems. Chaos: An Interdisciplinary Journal of Nonlinear Science. 2018;28(3):033611.

[7] Philippatos GC, Wilson CJ. Entropy, market risk, and the selection of efficient portfolios. Applied Economics. 1972;4(3):209–220.

[8] Fernholz ER. Stochastic portfolio theory. In: Stochastic Portfolio Theory. Springer; 2002. p. 1–24.

[9] Ou J. Theory of portfolio and risk based on incremental entropy. The Journal of Risk Finance. 2005;6(1):31–39.

[10] Xu J, Zhou X, Wu DD. Portfolio selection using $\lambda$ mean and hybrid entropy. Annals of Operations Research. 2011;185(1):213–229.

[11] Usta I, Kantar YM. Mean-variance-skewness-entropy measures: a multi-objective approach for portfolio selection. Entropy. 2011;13(1):117–133.

[12] Zhou R, Wang X, Dong X, Zong Z. Portfolio selection model with the measures of information entropy-incremental entropy-skewness. Advances in Information Sciences and Service Sciences. 2013;5(8):833.

[13] Zhang WG, Liu YJ, Xu WJ. A possibilistic mean-semivariance-entropy model for multi-period portfolio selection with transaction costs. European Journal of Operational Research. 2012;222(2):341–349.

[14] Bera AK, Park SY. Optimal portfolio diversification using the maximum entropy principle. Econometric Reviews. 2008;27(4-6):484–512.

[15] DeMiguel V, Garlappi L, Uppal R. Optimal Versus Naive Diversification: How Inefficient is the $1/N$ Portfolio Strategy? The Review of Financial Studies. 2009;22(5):1915–1953.

[16] Rödder W, Gartner IR, Rudolph S. An entropy-driven expert system shell applied to portfolio selection. Expert Systems with Applications. 2010;37(12):7509–7520.

[17] Chandrinos S, Lagaros ND. Construction Of Currency Portfolios By Means Of An Optimized Investment Strategy. Operations Research Perspectives. 2018;

[18] Gospodinov N, Maasoumi E, et al. General Aggregation of Misspecified Asset Pricing Models. Federal Reserve Bank of Atlanta; 2017.
[19] Chen X, Tian Y, Zhao R. Study of the cross-market effects of Brexit based on the improved symbolic transfer entropy GARCH model: An empirical analysis of stock–bond correlations. PLoS ONE. 2017;12(8):e0183194.

[20] Ormos M, Zibriczky D. Entropy-based financial asset pricing. PLoS ONE. 2014;9(12):e115742.

[21] Ponta L, Carbone A, Cincotti S. Detrending Moving Average algorithm: quantifying heterogeneity in financial data. In: Computer Software and Applications Conference (COMPSAC), 2017 IEEE 41st Annual. vol. 2. IEEE; 2017. p. 395–400.

[22] Ponta L, Carbone A. Information measure for financial time series: Quantifying short-term market heterogeneity. Physica A: Statistical Mechanics and its Applications. 2018;510:132–144. doi:https://doi.org/10.1016/j.physa.2018.06.085.

[23] Carbone A. Information Measure for Long-Range Correlated Sequences: the Case of the 24 Human Chromosomes. Scientific Reports. 2013;3:2721.

[24] Carbone A, Stanley HE. Scaling properties and entropy of long-range correlated time series. Physica A: Statistical Mechanics and its Applications. 2007;384(1):21–24.

[25] Carbone A, Castelli G, Stanley HE. Analysis of clusters formed by the moving average of a long-range correlated time series. Phys Rev E. 2004;69:026105.

[26] Hansen LP, Jagannathan R. Implications of Security Market Data for Models of Dynamic Economies. Journal of Political Economy. 1991;99(2):225–262.

[27] Hansen LP. Nobel Lecture: Uncertainty Outside and Inside Economic Models. Journal of Political Economy. 2014;122(5):945–987. doi:10.1086/678456.

[28] Hansen LP, Sargent TJ. Macroeconomic Uncertainty Prices. National Bureau of Economic Research; 2019. 25781.

[29] Backus D, Chernov M, Zin S. Sources of entropy in representative agent models. The Journal of Finance. 2014;69(1):51–99.

[30] Ghosh A, Julliard C, Taylor AP. What Is the Consumption-CAPM Missing? An Information-Theoretic Framework for the Analysis of Asset Pricing Models. Review of Financial Studies. 2017;30(2):442–504. doi:10.1093/rfs/hhw075.

[31] Shannon CE. A mathematical theory of communication, Part I, Part II. Bell Syst Tech J. 1948;27:623–656.
Table 1. Data series description. The analysis has been carried on tick-by-tick data of US indexes traded over the year 2018. Data time interval is about one second for all the three markets.

| Ticker  | Name                      | Country | Currency | Members | Length  |
|---------|---------------------------|---------|----------|---------|---------|
| NASDAQ  | NASDAQ COMPOSITE          | US      | USD      | 2570    | 6082017 |
| DJIA    | DOW JONES INDUS.          | US      | USD      | 30      | 5749145 |
| S&P500  | Standard & Poor 500       | US      | USD      | 505     | 6142443 |
Table 2. Data series lengths. The first column reports the number of periods in months $M$. The second, third and fourth column report the length of the price series in each period.

| $M$ | NASDAQ | DJIA | S&P500 |
|-----|---------|------|--------|
| 1   | 586866  | 516644| 516635 |
| 2   | 1117840 | 984101| 984046 |
| 3   | 1704706 | 1500764| 1500662|
| 4   | 2291572 | 1623779| 2017282|
| 5   | 2906384 | 2165044| 2558504|
| 6   | 3493250 | 2681708| 3075125|
| 7   | 4069315 | 3187571| 3580946|
| 8   | 4712062 | 3753440| 4146769|
| 9   | 5243029 | 4220774| 4614186|
| 10  | 5885781 | 4786624| 5180006|
| 11  | 6461845 | 5292487| 5685826|
| 12  | 6982017 | 5749145| 6142443|
Table 3. *Market dynamic index* $I(M, n)$ and *horizon dependence* $H(M, n)$ for the NASDAQ (top), DJIA (middle) and S&P500 (bottom) indexes. The spanned horizon ranges from one $M = 1$ to twelve $M = 12$. The values of the moving average window $n$ are reported in the first column. The reference value $I(1)$ has been calculated by assuming a consumption growth based on power utility, recursive utility and difference habit respectively in the third, fourth and fifth column following the approach [29].

|                  | Entropy | Power Utility | Recursive Utility | Difference Habit |
|------------------|---------|---------------|-------------------|------------------|
| **Nasdaq Composite Index (NASDAQ)** |         |               |                   |                  |
| 30 $I(12)$       | 0.0049  | 0.0214        | 0.0197            |                  |
| $H(12)$          | 0.0003  | 0.0012        | 0.0011            |                  |
| 50 $I(12)$       | 0.0052  | 0.0227        | 0.0209            |                  |
| $H(12)$          | 0.0003  | 0.0013        | 0.0012            |                  |
| 100 $I(12)$      | 0.0052  | 0.0229        | 0.0211            |                  |
| $H(12)$          | 0.0003  | 0.0015        | 0.0014            |                  |
| 150 $I(12)$      | 0.0054  | 0.0234        | 0.0215            |                  |
| $H(12)$          | 0.0005  | 0.0020        | 0.0018            |                  |
| 200 $I(12)$      | 0.0056  | 0.0246        | 0.0226            |                  |
| $H(12)$          | 0.0007  | 0.0032        | 0.0029            |                  |
| **Dow Jones Industrial Average Index (DJIA)** |         |               |                   |                  |
| 30 $I(12)$       | 0.0050  | 0.0218        | 0.0201            |                  |
| $H(12)$          | 0.0001  | 0.0004        | 0.0004            |                  |
| 50 $I(12)$       | 0.0050  | 0.0219        | 0.0201            |                  |
| $H(12)$          | 0.0001  | 0.0005        | 0.0004            |                  |
| 100 $I(12)$      | 0.0050  | 0.0217        | 0.0200            |                  |
| $H(12)$          | 0.0001  | 0.0003        | 0.0003            |                  |
| 150 $I(12)$      | 0.0050  | 0.0219        | 0.0201            |                  |
| $H(12)$          | 0.0001  | 0.0005        | 0.0004            |                  |
| 200 $I(12)$      | 0.0050  | 0.0218        | 0.0201            |                  |
| $H(12)$          | 0.0001  | 0.0004        | 0.0004            |                  |
| **S&P 500 Index (S&P500)** |         |               |                   |                  |
| 30 $I(12)$       | 0.0051  | 0.0224        | 0.0206            |                  |
| $H(12)$          | 0.0002  | 0.0010        | 0.0009            |                  |
| 50 $I(12)$       | 0.0052  | 0.0226        | 0.0208            |                  |
| $H(12)$          | 0.0003  | 0.0012        | 0.0011            |                  |
| 100 $I(12)$      | 0.0052  | 0.0227        | 0.0209            |                  |
| $H(12)$          | 0.0003  | 0.0013        | 0.0012            |                  |
| 150 $I(12)$      | 0.0052  | 0.0229        | 0.0211            |                  |
| $H(12)$          | 0.0003  | 0.0015        | 0.0014            |                  |
| 200 $I(12)$      | 0.0053  | 0.0230        | 0.0212            |                  |
| $H(12)$          | 0.0004  | 0.0016        | 0.0015            |                  |
Table 4. Relative entropy index and horizon dependence for representative agent models with constant variance (top), stochastic variance (middle) and jumps (bottom) as estimated by [29]

| Entropy | Power Utility | Recursive Utility | Ratio Habit | Difference Habit |
|---------|---------------|-------------------|-------------|------------------|
| $I(1) = EL_t(m_{t, t+1})$ | 0.0049 | 0.0214 | 0.0049 | 0.0197 |
| $I(\infty)$ | 0.0258 | 0.0232 | 0.0003 | 0.0258 |
| $H(120) = I(120) - I(1)$ | 0.0119 | 0.0011 | -0.0042 | 0.0001 |
| $H(\infty) = I(\infty) - I(1)$ | 0.0208 | 0.0018 | -0.0047 | 0.0061 |

Stochastic Variance

| Entropy | Recursive Utility 1 | Recursive Utility 2 | Campbell Cochrane | - |
|---------|---------------------|---------------------|-------------------|----|
| $I(1) = EL_t(m_{t, t+1})$ | 0.0218 | 0.0249 | 0.0230 | - |
| $I(\infty)$ | 0.0238 | 0.0293 | 0.0230 | - |
| $H(120) = I(120) - I(1)$ | 0.0012 | 0.0014 | 0 | - |
| $H(\infty) = I(\infty) - I(1)$ | 0.0020 | 0.0044 | 0 | - |

with Jumps

| Entropy | IID w/ Jumps | Stochastic Intensity | Constant Intensity 1 | Constant Intensity 2 |
|---------|---------------|----------------------|----------------------|----------------------|
| $I(1) = EL_t(m_{t, t+1})$ | 0.0485 | 0.0512 | 1.2299 | 0.0193 |
| $I(\infty)$ | 0.0485 | 0.0542 | 15.730 | 0.0200 |
| $H(120) = I(120) - I(1)$ | 0 | 0.0025 | 9.0900 | 0.0005 |
| $H(\infty) = I(\infty) - I(1)$ | 0 | 0.0030 | 14.5000 | 0.0007 |
Fig 1. Cluster entropy $S(\tau, n)$ vs duration $\tau$ for the time series of the prices (raw data) respectively of the market indices NASDAQ, DJIA and S&P500 described in Table 1. The series lengths are $N = 586866$, $N = 516644$ and $N = 516635$ respectively for NASDAQ, DJIA and S&P500 as given in Table 2. The curves refer to one period, i.e. the first month of tick-by-tick data ($M = 1$). Different curves in each figure refer to different values of the moving average window $n$ as indicated by the arrow.

Fig 2. Cluster entropy $S(\tau, n)$ vs duration $\tau$ for the time series of the prices (raw data) respectively of the market indices NASDAQ, DJIA and S&P500 described in Table 1. The series lengths are $N = 6982017$, $N = 5749145$ and $N = 6142443$ respectively for NASDAQ, DJIA and S&P500 as given in Table 2. The curves refer to twelve periods, i.e. the whole year 2018 of tick-by-tick data ($M = 12$). Different curves in each figure refer to different values of the moving average window $n$ as indicated by the arrow.
Fig 3. Cluster entropy $S(\tau, n)$ plotted versus cluster duration $\tau$ for the time series of the prices (sampled data) respectively of the market indices NASDAQ, DJIA and S&P500 described in Table 1. Figures refer to the first month of data ($M = 1$). All time series have same length $N = 492035$ obtained by a suitable sampling frequency. Different curves refer to different values of the moving average window $n$.

Fig 4. Cluster entropy $S(\tau, n)$ plotted versus cluster duration $\tau$ for the time series of the prices (sampled data) respectively of the market indices NASDAQ, DJIA and S&P500 described in Table 1. Figures refer to twelve months of data ($M = 12$). All time series have same length $N = 492035$ obtained by a suitable sampling frequency. Different curves refer to different values of the moving average window $n$. 
Fig 5. Market Dynamic Index $I(M, n)$ as a function of the moving average window $n$, calculated according to Eq. (11) for the prices respectively of the Nasdaq Composite, Dow Jones Industrial Average and S&P500 indexes as described in Table 2. Different curves in each figure refer to horizon varying from one ($M = 1$) to twelve months ($M = 12$). In particular, this set of curves corresponds to time series length $N = 492035$ with sampling frequency calculated as described Section 3. $I(M, n)$ has been evaluated as the integral of the entropy curves $S(\tau, n)$ similar to those shown in Figure 3. The insets show Market Dynamic Index $I(M, n)$ as a function of the horizon period unit $M$. Symbols with different colors refer to different values of the moving average window $n$ as indicated by the arrow (namely $n = 30s$, $n = 50s$, $n = 100s$, $n = 150s$ and $n = 200s$).
Supporting information

S1 Figure  Cluster entropy for artificially generated time series.

The artificial series has been generated with a total length equal to the NASDAQ index \((N = 6982017)\). Then the series has been divided in twelve consecutive segments with the same lengths of the NASDAQ subsequences (values of first column of Table 2). Figures refer respectively to the first segment \((M = 1)\), the sixth segment \((M = 6)\) and the twelveth segment \((M = 12)\) of the series. One can note that the cluster entropy is practically a constant in these cases.