Information measure for financial time series: Quantifying short-term market heterogeneity

Information measure for financial time series: quantifying short-term market heterogeneity

Linda Ponta[1] and Anna Carbone[2]

[1] *LIUC-Universitá Cattaneo, corso Giacomo Matteotti 22, 21053 Castellanza, Italy*

[2] *Politecnico di Torino, corso Duca degli Abruzzi 24, 10129 Torino, Italy*

Abstract

A well-interpretable measure of information has been recently proposed based on a partition obtained by intersecting a random sequence with its moving average. The partition yields disjoint sets of the sequence, which are then ranked according to their size to form a probability distribution function and finally fed in the expression of the Shannon entropy. In this work, such entropy measure is implemented on the time series of prices and volatilities of six financial markets. The analysis has been performed, on tick-by-tick data sampled every minute for six years of data from 1999 to 2004, for a broad range of moving average windows and volatility horizons. The study shows that the entropy of the volatility series depends on the individual market, while the entropy of the price series is practically invariant for the six markets. Finally, a cumulative information measure - the *Market Heterogeneity Index* - derived from the integral of the entropy measure, is introduced for obtaining the weights of an *Efficient Portfolio*. A comparison with the weights obtained by using the Sharpe ratio - a traditional risk diversity measure - is also reported.

Keywords: Entropy; Long-range correlated time series, Market Heterogeneity, Portfolio Selection

1. Introduction

Several connections between economics and statistical thermodynamics have been suggested over the years. Marginal utility and disutility have been related respectively to force, energy and work [1]. Differential pressure and volume in an ideal gas have been linked to price and volume in financial systems [2], money utility has been linked to entropy [3] and temperature has been linked to velocity of circulation of money [4]. As a result of the growing diversification and globalization of economy, applications of entropy concepts to finance and economics are receiving renewed attention as instruments to monitor and quantify market diversity [5–17]. Heterogeneity of private and institutional investments might result, among other effects, in imperfect and asymmetric flow of information at a microscopic level, ultimately undermining the validity of the efficient market theory with its assumption of a homogeneous random process underlying the stock prices dynamics [18].

The interplay of noise and profitability with specific focus on volatility investment strategies has therefore gained interest and is under intense scrutiny [19–22]. Volatility series exhibit remarkable features suggesting the existence of some amount of 'order' out of the seemingly random structure. The degree of 'order' is intrinsically linked to the information, embedded in

Preprint submitted to Elsevier *April 28, 2020*

the volatility patterns, whose extraction and quantification might shed light on microscopic financial phenomena [8–15]. The Volatility Index (VIX), defined as the near-term volatility conveyed by $S\&P500$ stock index option prices, has been suggested as a figure able to monitor investor sentiment. VIX futures and options have been introduced as trading instruments. The Chicago Board Options Exchange (`http://www.cboe.com/`) calculates and updates the values of more than 25 indexes designed to measure the expected volatility of different securities.

New tools for portfolio optimization and asset pricing should be based on the expected volatility over different horizons to evaluate market performances beyond the figures of risk provided by the traditional static techniques [23–25]. Markowitz portfolio selection model (also known as 'mean-variance' model) and its variants are based on the assumption that returns of assets follow a normal distribution, a condition that has been extensively proved to be far from the real behaviour of financial data. Portfolios highly concentrated on a limited number of assets are provided by the Markowitz method, which therefore do not fulfil the purpose of optimally diversifying the investment. Another drawback of the 'mean-variance' model is the poor resolution of data outliers. For assets that are asymmetrically or non-normally distributed, improved measures of uncertainty have been proposed: dynamic rather than static; based on moments higher than the first/second (mean/variance); not relying on a specific distribution of prices/returns.

Entropy, as a quantitative estimate of diversity, might prove convenient for portfolio selection. The entropy concept, exploited for portfolio selection and compared to traditionally estimates of *optimal portfolios* in [26–28], was found consistent with the Markowitz full-covariance and the Sharpe single-index models. The incremental entropy approach has been proposed in [29] where in particular the need of building an optimal portfolio for a given probability of return was remarked. Entropy based optimal portfolios, where the asset risk are caused by randomness and fuzziness, have been put forward in [30–32]. The mean-variance-skewness-entropy model, proved better than traditional portfolio for out-of-sample tests [33–35]. A possibilistic mean-semivariance-entropy model for multi-period portfolio selection with transaction costs is proposed in [36]. The portfolio weights by a rule-based inference mechanism under both maximum entropy and minimum relative entropy [37–42].

In this work, the information measure approach proposed in [43–45] is applied to prices and volatilities of tick-by-tick data, recorded from 1999 to 2004, of six financial markets. The method here proposed does not suffer of the 'mean-variance' model limitations as it does not require a specific distribution of returns, i.e. in particular a symmetric Gaussian distribution. The investigation is performed over a broad range of volatility horizons and moving average windows. Interestingly, it is found that the entropy of the volatility series takes different values for the different markets as opposed to the entropy of the prices, which is mostly constant and consistent with a homogeneous random walk structure of the six market time series. In order to provide a clear procedure for the practical use of the proposed approach, a synthetic figure - Market Heterogeneity Index - is introduced, defined in terms of the integral of the entropy functional. Hence, the Market Heterogeneity Index is calculated for prices and volatilities of the six financial markets, over different volatility horizons and moving average windows. The values of the index are finally compared with the results obtained by using the Sharpe ratio, a traditional estimate of portfolio risk.

Different approaches have been proposed for the evaluation of the entropy of a random sequence $\{x_t\}$, whose preliminary fundamental step is the symbolic representation of the data, through a partition suitable to map the continuous phase-space into disjoint sets [46–52]. The method commonly adopted for partitioning a sequence is based on a uniform division in equal size blocks. Then the entropy is estimated over subsequent partitions corresponding to different block sizes [46]. The construction of an optimal partition is not a trivial task, as it is crucial to effectively discriminate between randomness/determinism of the encoded/decoded data (see Section III of Ref.[53] for a short review of the partition methods).

Here, the partition is obtained by taking the intersection of $\{x_t\}$ with the moving average $\{\widetilde{x}_{t,n}\}$ for different moving average window n [43–45]. For each window n, the subsets $\{x_t : t = s, ..., s-n\}$ between two consecutive intersections are ranked according to their size to obtain the probability distribution function P. The elements of these subsets are the segments between consecutive intersections and have been named *clusters* (see the illustration shown in Fig. 1). The present approach directly yields either power-law or exponential distributed blocks (clusters), thus enabling us to separate the sets of inherently correlated/uncorrelated blocks along the sequence. Moreover, 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. For the sake of clarity, the main relationships relevant to the investigation carried in this paper will be shortly recalled in Section 2.

2. Cluster entropy

The approach adopted in the framework of this work [43–45], stem from the idea of Claude Shannon to quantify the expected information contained in a message extracted from a sequence x_t [54] by using the entropy functional:

$$S[P] = \sum_j p_j \log p_j \ . \tag{1}$$

with P a probability distribution function associated with the time sequence x_t.

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

$$\tau_j \equiv \|t_j - t_{j-1}\| \tag{2}$$

where the instance t_{j-1} and t_j refer to two subsequent intersections as shown in Fig.1.

The probability distribution function $P(\tau, n)$ can be obtained by ranking the number of clusters $\mathcal{N}(\tau_1, n), \mathcal{N}(\tau_2, n), ..., \mathcal{N}(\tau_j, n)$ according to their length $\tau_1, \tau_2, ..., \tau_j$ for each n.

For a fractional Brownian motion, a stationary sequence of self-affine clusters C is generated with probability distribution function varying as [43]:

$$P(\tau, n) \sim \tau^{-\alpha}\mathcal{F}(\tau, n) \ , \tag{3}$$

with the factor $\mathcal{F}(\tau, n)$ taking the form $\exp(-\tau/n)$, to account for the finite size effects when $\tau \gg n$, resulting in the drop-off of the power-law and the onset of the exponential decay.

By using Eq. (3), Eq. (1) writes (the details of the derivation can be found in [43, 45]):

$$S(\tau, n) = S_0 + \log \tau^{\alpha} + \frac{\tau}{n} \ , \tag{4}$$

where S_0 is a constant, $\log \tau^{\alpha}$ and τ/n are related respectively to the terms $\tau^{-\alpha}$ and $\mathcal{F}(\tau, n)$.

The constant S_0 in Eq. (4) can be evaluated as follows: in the limit $n \sim \tau \to 1$, $S_0 \to -1$ consistently with the minimum value of the entropy $S(\tau, n) \to 0$ that corresponds to a fully ordered (deterministic) set of clusters with same duration $\tau = 1$. The maximum value of the entropy $S(\tau, n) = \log N^\alpha$ is then obtained when $n \sim \tau \to N$ with N the maximum length of the sequence. This condition corresponds to the maximum randomness (minimum information) carried by the sequence, when a single cluster is obtained coinciding with the whole series.

The term $\log \tau^\alpha$ in Eq. (4) can be written as $\log \tau^D$, since the exponent α is equal to the fractal dimension $D = 2 - H$ with H the Hurst exponent of the time series. The term $\log \tau^D$ can be thus interpreted as a generalized form of the Boltzmann entropy $S = \log \Omega$, where $\Omega = \tau^D$ can be thought of the volume occupied by the fractional random walker.

The term τ/n in Eq. (4) 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. This aspect has been discussed for this specific measure in relation to the DNA sequences in [43], while the general issue of the finite size effects on entropy measure has been discussed in Refs. [55–58]

The method and Eqs. (1 - 4) have been applied for estimating the probability distribution function $P(\ell, n)$ and the entropy $S(\ell, n)$ of the cluster length ℓ of the 24 nucleotide sequences of the human chromosomes in [43]. It is worth noting that the characteristic sizes ℓ and τ (respectively cluster length and duration) are equivalent from a statistical point of view. However, the cluster duration τ is a variable more suitable and meaningful than the cluster length ℓ when dealing with time series $\{x_t\}$. Furthermore, the cluster duration τ is particularly relevant to financial market series, due to its straightforward connection to the duration of the investment horizon via the volatility window T and the moving average window n, entering as parameters in the above calculation.

Before entering the details of the application to the six financial indexes, it is worthy to clarify the implication of partitioning a random sequence either by using boxes having equal size or by using the clusters obtained through the intersection with the moving average. For same size boxes, the excess noise term τ/n vanishes, thus the entropy reduces to the logarithmic term (see Eq. (8) in Ref. [46]) which corresponds to the intrinsic entropy of an ideal fractional random walk of dimension $D = 2 - H$. When a moving average partition is used, an excess entropy term τ/n emerges accounting for the additional heterogeneity introduced by the random partitioning process operated by using the moving average intersections. This is particular relevant for financial applications as it represents a measure of the disorder introduced by the trading mechanism based on moving averages.

3. Data

The information measure described in the previous section is here applied to financial prices and volatilities of six market indices (BOBL, BUND, DAX, Euro Currency, Euro Stoxx and FIB30). For each index, the data set includes tick-by-tick prices p_t sampled every minute from 4 January 1999 to 24 March 2004. The length of the series is set for all the six markets equal to 517041, determined on the basis of the length of the shortest series which is the Eurocurrency. The six time series are shown in Fig. 2.

Among other reasons, the main motivation for selecting the six markets mentioned above is the fact that they operate in a very similar socio-economic context, being all traded within the EU zone. This rules out that the diversity featured by the entropy measure might stem from exogenous drives rather than from the intrinsic dynamics of the market. A short description of the data and the markets is provided here below.

BUND is a debt security issued by Germany's federal government, and it is the German equivalent of a U.S. Treasury bond. The German government uses bunds to finance its spending, and bonds with long-term durations are the most widely issued securities. Bunds are auctioned only with original maturities of 10 and 30 years. Bunds represent long-term obligations of the German federal government that are auctioned off in the primary market and traded in the secondary market. Bunds can be stripped, meaning their coupon payments can be separated from their principal repayments and traded individually. Bunds pay interest and principal typically once a year, and they represent an important source of financing for the German government. The principal characteristics of bunds are that they are nominal bonds with fixed maturities and fixed interest rates. All German government debt instruments, including bunds, are issued by making a claim in the government debt register rather than producing paper certificates. A typical bund issue will state its issuance volume, maturity date, coupon rate, payable terms and interest calculation standard used.

BOBL is a futures contract with medium term debt that is issued by the Federal Republic of Germany as its underlying asset. The contract has a notional contract value of 100,000 euros, with a term to maturity of 4.5 to five years. Unlike most other types of future contracts, BOBL future contracts tend to be settled by delivery. In America, these futures contracts are traded on the Chicago Board of Trade, under the symbol GBM. The Euro-Bobl future, along with the Euro-Bund and Euro-Schatz futures, are the most heavily traded fixed-income securities in the world. BOBL is an acronym for bundesobligationen. This translates to 'federal obligations' in English .

DAX, the "Deutscher Aktien indeX", is the derivative of the main German stock index consisting of the 30 major German companies trading on the Frankfurt Stock Exchange. Just like the FTSE 100 and S&P500, DAX is a capitalization-weighted index so it essentially measures the performance of the 30 largest, publicly traded companies in Germany. It is therefore a strong indicator of the strength of the German economy and investor sentiment towards German equities. The DAX has been a relatively stable index with 16 companies of the original 30 remaining in the index since its inception in 1988. The index began with a base date of 30 December 1987 and a base value of 1,000. Over the years the DAX has seen a large amount of takeovers, mergers, bankruptcies, and restructurings. Since 1 January 2006, the index is calculated after every seconds. It is computed daily between 09:00 and 17:30 CET.

Euro Stoxx is an index of the Eurozone stocks designed by STOXX, a provider owned by Deutsche Brse Group. The Euro Stoxx 50 provides a blue-chip representation of Supersector leaders in the Eurozone. It is made up of fifty of the largest and most liquid stocks. The Euro Stoxx 50 was introduced on 26 February 1998. Its composition is reviewed annually in September. The index is available in several currency (EUR, USD, CAD, GBP, JPY) and return (Price, Net Return, Gross Return) variant combinations. Calculation takes place every 15 seconds between 09:00 CET and 18:00 CET for the EUR and USD variants of any return type, while the CAD, GBP and JPY variants are available as end-of-day calculation only (18:00 CET). The Euro Stoxx 50 Index is derived from the 19 Euro Stoxx regional Supersector indices and represents the largest super-sector leaders in the Eurozone in terms of free-float market capitalization. The index captures about 60% of the free-float market capitalization of the Euro Stoxx Total Market

Index (TMI), which in turn covers about 95% of the free-float market capitalizaion of the represented countries. It is one of the most liquid indices for the Eurozone: an ideal underlying for financial products or for benchmarking purposes. Additionally, the index serves as an underlying for many strategy indices, such as the Euro Stoxx 50 Risk Control Indices. Buffers are used to achieve the fixed number of components and to maintain stability of the indices by reducing index composition changes. Selection methodology ensures a stable and up-to-date index composition. Fast-entry and fast-exit rules ensure the index accurately represents the performance of only the biggest and most liquid stocks.

Euro Currency Index (EUR_I) represents the arithmetic ratio of four major currencies against the Euro: US-Dollar, British Pound, Japanese Yen and Swiss Franc. All currencies are expressed in units of currency per Euro. The index was launched in 2004 by the exchange portal Stooq.com. Underlying are 100 points on 4 January 1971. Based on the progression, Euro Currency Index can show the strength or weakness of the Euro. A rising index indicates an appreciation of the Euro against the currencies in the currency basket, a falling index in contrast, a devaluation. Relationships to commodity indices are recognizable. A rising Euro Currency Index means a tendency of falling commodity prices. This is especially true for agricultural commodities and the price of oil. Even the prices of precious metals (gold and silver) are correlated with the index. Arithmetically weighted Euro Currency Index is comparable to the trade-weighted Euro Effective exchange rate index of the European Central Bank (ECB). The index of ECB measures much more accurately the value of the Euro, compared to the Euro Currency Index, since the competitiveness of European goods in comparison to other countries and trading partners is included in it.

FIB30 is the derivative of the MIB 30 stock market index consisting of the 30 major Italian companies trading on the Borsa Italiana until the 2003. In fact in 2003 it is substituted by the S&P/MIB, and in the 2009 by the FTSE-MIB index based on the 40 major company traded in the Italian market. The FTSE MIB index is the major benchmark index of the Italian stock markets. This index, which accounts for about 80% of the domestic market capitalization, consists of companies of primary importance and high liquidity in the various ICB sectors in Italy. The FTSE MIB Index measures the performance of 40 Italian securities and intends to reproduce the weights of the expanded Italian equity market. The Index is derived from the trading universe of securities on the main stock market of Borsa Italiana (BIt). Each title is analyzed by size and liquidity, and the Index provides a representation of sectors. The FTSE MIB is a weighted index based on market capitalization. It is computed daily between 9:00 and 17:50 CET.

4. Results

In this work, the continuously compounded return:

$$r_t = p_t - p_{t-h} \quad , \tag{5}$$

and the log-return:

$$r_t = \log p_t - \log p_{t-h} \quad , \tag{6}$$

have been considered, with $0 < h < t < N$ and N the maximum length of the time series. The plots of the prices, the linear return and logarithmic return for the six markets are shown in Fig. 2, Fig. 3 and Fig. 4 respectively.

The volatility $v_{t,T}$ has been taken as:

$$v_{t,T} = \sqrt{\frac{\sum_{t=1}^{T} (r_t - \mu_T)^2}{T-1}} \quad , \tag{7}$$

where r_t are the above defined return, T the volatility window, $\mu_T = \sum_{t=1}^{T} r_t/(T-1)$ the average value of the return taken over the window T. It is well known that the volatility definition is not univocal, however this work has been performed by using the definition of Eq. (7).

The series of the prices of the six markets, shown in Fig. 2, are used to calculate the probability distribution functions $P(\tau, n)$ and the entropy $S(\tau, n)$ according to the above described procedure. The probability $P(\tau, n)$ for the six price series is shown in Fig. 5. The analysis has been performed for n ranging between 5 min and 1500 min, however for the clarity of visualization, only the curves corresponding to moving average windows $n = 30$ min, $n = 50$ min, $n = 100$ min, $n = 150$ min, $n = 200$ min are plotted. The curves exhibit a power law behaviour for $\tau < n$, and an exponential decay, for $\tau > n$, respectively. The power-law exponents agree with the theoretical prediction $\alpha = 2 - H$ expected for fractional Brownian motions. The curves shown in Fig. 5 are thus consistent with Eq. (3) over the whole range of investigated values.

Fig. 6 shows the entropy $S(\tau, n)$ calculated by using the probability distribution functions plotted in Fig. 5. The curves correspond to moving average values $n = 30$ min, $n = 50$ min, $n = 100$ min, $n = 150$ min, $n = 200$ min. The range of investigated moving average windows is much broader (from $n = 5$ min to $n = 1500$ min) but it is not shown for visualization clarity. The behavior of the curves is quite carefully reproduced by Eq. (4): the first part increases as a logarithmic function, then for $\tau \approx n$ a sharp increase is observed corresponding to the onset of the linear term τ/n. One can note that $S(\tau, n)$ is n-invariant for small values of τ, while its slope decreases as $1/n$ at larger τ, as expected according to Eq. (4), 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 τ can be generated by different values of the moving average window n. For constant τ, larger entropy values are obtained as n increases.

The probability distribution function $P(\tau, n)$ and the entropy $S(\tau, n)$ have been calculated for a large set of volatility series with (i) linear and (ii) logarithmic return with volatility window T ranging from half a trading day to 20 trading days (namely $T = 330$ min, $T = 660$ min, $T = 1320$ min, $T = 1980$ min, $T = 2640$ min, $T = 3300$ min, $T = 3960$ min, $T = 4620$min, $T = 5280$ min, $T = 5940$ min, $T = 6600$ min, $T = 13200$ min). The entropy $S(\tau, n)$ for the volatility of logreturn series with $T = 660$ min is shown in Fig. 7 for the six markets.

By comparing the results shown in Fig. 6 and in Fig. 7, one can note a general increase of the entropy of the volatilities compared to prices. Such an increase has been systematically found in all the analysed series, both for linear and logarithmic return with diverse volatilities windows T. Preliminary results can be found in [59] where the invariance of the entropy curves for the prices is shown. The volatility entropy behaviour is related to a deviation of the probability distribution function of real-world time series due to the reduced randomness and data redundancy compared to the ideal one predicted on the basis of Eq. (3) deduced for fractional Brownian motions. These issues are further enphasized when the data are fed in the volatility relationship Eq. (7) which, being a variance, enhances the variability compared to that of the prices.

5. Portfolio Selection

On the basis of the findings reported in the previous section, we put forward a procedure to estimate an information heterogeneity index from the entropy of the volatility series. Then we also show how to use this index for estimating the weights of portfolio assets. Before discussing the market heterogeneity indexes obtained from the cluster entropy data, the main steps of the Markowitz approach are briefly recalled for the sake of completeness.

When a rational investor should compare two stocks with the same return, the choice goes to the one with the smaller variance (lower risk). A generalized reasoning was pioneered by Markowitz to portfolio selection with more than two stocks. The expected return for an portfolio is defined by Markowitz as:

$$\overline{r}_P = E\left[\sum_{k=1}^{M} w_k r_{k,t}\right] = \sum_{k=1}^{M} w_k \mu_k \ , \tag{8}$$

where $r_{k,t}$ are the return defined by Eqs. (5,6) and w_k is the percentage of the investment for each stock k. The portfolio variance is defined as:

$$\sigma_P^2 = \sum_{k=1}^{M}\sum_{l=1}^{M} w_k \sigma_{k,l} w_l \ , \tag{9}$$

where:

$$\sigma_{k,l} = E\left[(r_{k,t} - \mu_{k,t})(r_{l,t} - \mu_{l,t})\right] \ , \tag{10}$$

with k and l any two out of the M stocks.

The goal of the portfolio choice problem is to seek minimum risk for a given level of return and to seek maximum return for a given level of risk. In formulas, the portfolio optimization problem minimizes:

$$\frac{1}{2}\sigma_P^2 - \frac{1}{\lambda}\overline{r}_P \ , \tag{11}$$

with the basic portfolio constraints:

1. $w_k \geq 0$ (the portfolio weights are set to be nonnegative by using the lower-bound constraint);
2. $\sum_{k=1}^{M} w_k = 1$ (the portfolio weights sum is set to 1 by using the budget constraint).

Portfolios satisfying these criteria are called *efficient* and the graph of the risks and returns of these portfolios forms a curve called the *efficient frontier* [23]. The portfolio that maximizes the Sharpe ratio is among the various optimal portfolios on the efficient frontier, the one identified by the intersection between the tangent line from the risk-free rate to the efficient frontier and the efficient frontier. The Sharpe ratio is defined as the average return earned in excess of the risk-free rate r_f per unit of volatility or total risk [24, 25]:

$$\mathcal{R}_S = \frac{\overline{r}_P - r_f}{\sigma_P} \ , \tag{12}$$

where $\overline{r}_p$ is the expected portfolio return and σ_p the portfolio standard deviation, and r_f is the risk free rate defined as:

$$\overline{r}_f = E\left[\sum_{k=1}^{M} r_{k,t}\right] = \sum_{k=1}^{M} \mu_k \ . \tag{13}$$

Generally, the greater the value of the Sharpe ratio, the more attractive the risk-adjusted return.

As extensively reported in the literature, the Markowitz approach suffers from many drawbacks: (1) it is based on the unrealistic assumption of Gaussian distributed returns; (2) it provides portfolios that are highly concentrated on few highly concentrated assets, which does not correspond to the maximal diversification. In order to reduce the effects of this limitation, the entropy concept has been applied to the vector of weights w_k. The portfolio weights are maximally diversified when the entropy is maximal, i.e. when all the states are equally likely $\{w_k = 1/M\}_{k=1,...,M}$. The entropy is minimal in the deterministic case $\{w_k = 1\}_{k=k'}$ otherwise $\{w_k = 0\}_{k\neq k'}$. In practice, first the weights are estimated by using the above equations and then the entropy of the weights w_k is maximized by using the Shannon entropy [12, 16, 37, 38].

The approach proposed in this work takes a different perspective as discussed in the following paragraphs. Here, the entropy estimate is straightaway obtained from the financial market data without using the mean-variance hypothesis of Gaussian returns.

To univocally quantify the entropy behaviour of volatilities of each single market, we introduce a *cumulative information measure* defined as follows:

$$I_{k,C}(n) = -\int_0^{\tau_{max}} S_k(\tau, n) d\tau \quad . \tag{14}$$

where the index k refers to each single market. The function $I_{k,C}(n)$ has been plotted in Fig. 8 for the prices (a) and the volatilities with linear (b) and logarithmic return (c). The function $I_{k,C}(n)$ is almost constant for the prices of the six markets (a), conversely it takes different values for the volatilities (b,c).

The quantity $I_{k,C}(n)$ is further integrated over n:

$$I_{k,C} = \int_{n_{min}}^{n_{max}} I_{k,C}(n) dn \quad . \tag{15}$$

The quantity $I_{k,C}$ given by Eq. (15) is a cumulative figure of heterogeneity, a number suitable to quantify the reduced information content (excess entropy) and make a comparison across the different markets. The absolute value of the Market Heterogeneity Index $I_{k,C}$ has been plotted in Fig. 9 for the six markets.

For each market k and volatility windows T, the value of $I_{k,C}$ calculated according to Eq. (15) has been rescaled between 0 and 1 using:

$$\overline{I}_{k,C} = \frac{\frac{I_{k,C}}{\sum_{k=1}^{M} I_{k,C}} - \min\left[I_{k,C}\right]}{\max\left[I_{k,C}\right] - \min\left[I_{k,C}\right]} \tag{16}$$

where M is the set of all markets, in the present work $M = 6$. In particular the data in Fig. 9 (a) correspond to the integrals under the curves shown in Fig. 8 (a) for the prices. The data of Fig. 9 (b) correspond to the integrals of the volatility curves with linear return shown in Fig. 8 (b). The data of Fig. 9 (c) correspond to the integrals of the volatility curves with the logarithmic return shown in Fig. 8 (c). Different symbols corresponds to the different volatility windows, namely $T = 1320$ min (circles), $T = 1980$ min (squares), $T = 2640$ min (up triangles), $T = 3300$ min (down triangles). Consistently with the results shown in Fig. 8, a strong variability is observed in the volatility curves (b,c) compared to prices (a). Finally , the fraction of investments in each market (portfolio weights) normalized to 1 can be obtained by taking the following quantities:

$$w_k = 1 - \overline{I}_{k,C} \quad . \tag{17}$$

In order to reduce the spurious fluctuations due to the different markets opening/closing and improve the quality of the estimate, an average value of $\overline{I}_{k,C}$ might be taken over the set of volatility windows T. The values of the weights are shown in Fig. 10 (up triangles) estimated by using the data plotted Fig. 9 (b). The weights of the efficient portfolio that maximizes the Sharpe Ratio are shown for comparison in Fig. 10 (circles). The portfolio weights obtained via the market heterogeneity strategy and the Sharpe ratio are also shown in Table 1.

The Sharpe ratio values reported in Table 1 have been obtained by using the portfolio optimization tool provided by MATLAB financial toolbox [60]. The main steps of the portfolio optimization tool are the following:

- Create a portfolio object by using the ‘Portfolio’ function.
- Estimate mean and covariance of the return. By using the ‘setAssetMoments’ function and its properties ‘AssetMean’ (for the mean) and ‘AssetCovar’ (for the covariance) are then assigned to the ‘Portfolio’ object.
- Specify the ‘Portfolio’ constraints. The basic constraints for the optimization, given by Eqs. (11), are set. Additional constraints such as linear equality and inequality, bound, budget, group, group ratio, turnover, and tracking error might also be considered.
- Validate the ‘Portfolio’ for identifying errors by using specialized validation functions.
- Estimate the *efficient portfolios* and *efficient frontier*. The ‘estimateFrontier’ function computes efficient portfolios spaced evenly according to the return proxy from the minimum to maximum return efficient portfolios. The number of portfolios estimated is controlled by the hidden property ‘defaultNumPorts’ which is set to 10 in the present analysis.
- Choose the portfolio that maximizes the Sharpe ratio. By using the ‘estimateMaxSharpeRatio’ function, the optimal portfolio that maximizes the Sharpe ratio among the portfolios on the efficient frontier is identified. This function implements Eq. (12), where the expected portfolio return is given by Eq. (8), the portfolio variance by Eq. (9) and the risk free rate, if not exogenously given, is set by default equal to zero.

Looking at Table 1 there is a very good correspondence between the weights obtained with the two strategies. It is worth noting that the portfolio that maximizes the Sharpe ratio correspond to a strategy limited to only three out of the six markets, as opposed to the cluster entropy strategy that has a smoother and better distributed variation of the weights. It is worthy of note that the investment horizon of the cluster entropy measure is defined by the range of the duration τ, then it is of the order or smaller than the maximum value of the moving average window n. Conversely, the optimization based on the Sharpe ratio refers to the whole duration of the financial data series. Therefore the time scale of the proposed measure refers to a short term investment horizon (day-by-day trading).

Conclusion

The implementation of the moving average algorithm to estimate the Shannon entropy of a long-range correlated sequence has been illustrated to analyse the tick-by-tick data of six markets BOBL, BUND, DAX, Euro Currency, Euro Stoxx and FIB30 from 1999 to 2004. By considering several runs of the algorithm for the different parameters (mainly moving average n and volatility

Table 1
Portfolio weights obtained by using the Sharpe ratio (1^{st} row) and the Cluster Entropy (2^{nd} row) for the six markets. The weights have been obtained from the market heterogeneity indexes shown in Fig. 9 (b). Similar values are obtained by using the market heterogeneity indexes with logreturn shown Fig. 9 (c). It is worthy of note that the investment horizon of the current entropy measure is defined by the range of the cluster duration τ i.e. of the order or smaller than the maximum moving average window n. Conversely, the optimization based on the Sharpe ratio refers to the whole duration of the financial data series. Therefore the comparison between the two strategies makes sense only at a qualitative level.

Portfolio weights	BOBL	BUND	DAX	Euro Stoxx	Euro Currency	FIB30
w_k (Sharpe ratio)	0.6189	0.3566	0	0	0.0245	0
w_k (Cluster entropy)	0.2007	0.2218	0.1913	0.1689	0.1471	0.0702

T windows) this study has systematically shown, the entropy measured on the price is practically market invariant, whereas the entropy of the volatilities is market dependent. A cumulative market heterogeneity index, defined by Eqs. (15 -17) has been built and its values compared with those of a standard risk measure (the Sharpe ratio). Compared to the Sharpe ratio, the market heterogeneity index provides a more accurate and smooth evaluation of the portfolio composition, that might be an advantage for practical applications of the entropy measure. The novelty of the present approach resides in the method used for partitioning the sequence which allows one to separate the sets of inherently informative/uninformative clusters along the sequence. Moreover, the clusters are exactly defined as the portions of the series between death/golden crosses according to the technical traders rules. Therefore, the information content quantified by this approach is intimately linked to technical trader's viewpoint on the analysed markets. The six markets used in this study have been chosen because they operate in close socio-economic contexts within the EU. The similarity of the ambient context permits to rule out external causes as the source of the market heterogeneity, while ensuring that the diversity is due to endogenous processes. The current work will be extended to evaluate portfolios including markets beyond Europe and other more volatile assets.

References

[1] I. Fisher, Mathematical investigations in the theory of value and prices, Yale University.
[2] P. A. Samuelson, Maximum principles in analytical economics: Nobel lecture 1970, The American Economic Review 62 (3) (1972) 249–262.
[3] J. H. Lisman, Econometrics and thermodynamics: a remark on davis' theory of budgets, Econometrica: Journal of the Econometric Society (1949) 59–62.
[4] A. G. Pikler, Utility theories in field physics and mathematical economics (ii), The British Journal for the Philosophy of Science 5 (20) (1955) 303–318.
[5] J. Chen, The unity of science and economics: A new foundation of economic theory, Springer, 2015.
[6] J. Le Gallo, S. Dall Erba, Evaluating the temporal and spatial heterogeneity of the european convergence process, 1980–1999, Journal of Regional Science 46 (2) (2006) 269–288.
[7] L. Ponta, S. Pastore, S. Cincotti, Information-based multi-assets artificial stock market with heterogeneous agents, Nonlinear Analysis: Real World Applications 12 (2) (2011) 1235 – 1242.
[8] R. Zhou, R. Cai, G. Tong, Applications of entropy in finance: A review, Entropy 15 (11) (2013) 4909–4931.
[9] M. Ormos, D. Zibriczky, Entropy-based financial asset pricing, PLoS ONE 9 (12) (2014) e115742.
[10] J.-R. Yu, W.-Y. Lee, W.-J. P. Chiou, Diversified portfolios with different entropy measures, Applied Mathematics and Computation 241 (2014) 47–63.
[11] M. Sheraz, S. Dedu, V. Preda, Entropy measures for assessing volatile markets, Procedia Economics and Finance 22 (2015) 655–662.
[12] G. Pola, On entropy and portfolio diversification, Journal of Asset Management 17 (4) (2016) 218–228.

[13] J. Contreras, Y. E. Rodríguez, A. Sosa, Construction of an efficient portfolio of power purchase decisions based on risk-diversification tradeoff, Energy Economics 64 (2017) 286–297.
[14] S. Bekiros, D. K. Nguyen, S. J. L., G. S. Uddin, Information diffusion, cluster formation and entropy-based network dynamics in equity and commodity markets, European Journal of Operational Research 256 (3) (2017) 945–961.
[15] X. Chen, Y. Tian, R. Zhao, Study of the cross-market effects of brexit based on the improved symbolic transfer entropy garch modelan empirical analysis of stock–bond correlations, PLoS ONE 12 (8) (2017) e0183194.
[16] A. Meucci, Risk and asset allocation, Springer Science & Business Media, 2009.
[17] A. Meucci, M. Nicolosi, Dynamic portfolio management with views at multiple horizons, Applied Mathematics and Computation 274 (2016) 495–518.
[18] E. F. Fama, Efficient capital markets: A review of theory and empirical work, The Journal of Finance 25 (2) (1970) 383–417.
[19] A. Gunasekarage, D. M. Power, The profitability of moving average trading rules in south asian stock markets, Emerging Markets Review 2 (1) (2001) 17–33.
[20] L. Menkhoff, M. P. Taylor, The obstinate passion of foreign exchange professionals: technical analysis, Journal of Economic Literature 45 (4) (2007) 936–972.
[21] M. Frömmel, K. Lampaert, Does frequency matter for intraday technical trading?, Finance Research Letters 18 (2016) 177–183.
[22] D. M. Smith, N. Wang, Y. Wang, E. J. Zychowicz, Sentiment and the effectiveness of technical analysis: Evidence from the hedge fund industry, Journal of Financial and Quantitative Analysis 51 (6) (2016) 1991–2013.
[23] H. Markowitz, Portfolio selection, The Journal of Finance 7 (1) (1952) 77–91.
[24] W. F. Sharpe, Capital asset prices: A theory of market equilibrium under conditions of risk, The Journal of Finance 19 (3) (1964) 425–442.
[25] W. F. Sharpe, Mutual fund performance, The Journal of Business 39 (1) (1966) 119–138.
[26] G. C. Philippatos, C. J. Wilson, Entropy, market risk, and the selection of efficient portfolios, Applied Economics 4 (3) (1972) 209–220.
[27] A. K. Bera, Y. Bilias, The mm, me, ml, el, ef and gmm approaches to estimation: a synthesis, Journal of Econometrics 107 (1-2) (2002) 51–86.
[28] E. R. Fernholz, Stochastic portfolio theory, in: Stochastic Portfolio Theory, Springer, 2002, pp. 1–24.
[29] J. Ou, Theory of portfolio and risk based on incremental entropy, The Journal of Risk Finance 6 (1) (2005) 31–39.
[30] K. Smimou, C. Bector, G. Jacoby, A subjective assessment of approximate probabilities with a portfolio application, Research in International Business and Finance 21 (2) (2007) 134–160.
[31] X. Huang, Mean-entropy models for fuzzy portfolio selection, IEEE Transactions on Fuzzy Systems 16 (4) (2008) 1096–1101.
[32] J. Xu, X. Zhou, D. D. Wu, Portfolio selection using λ mean and hybrid entropy, Annals of Operations Research 185 (1) (2011) 213–229.
[33] P. Jana, T. Roy, S. Mazumder, Multi-objective possibilistic model for portfolio selection with transaction cost, Journal of Computational and Applied Mathematics 228 (1) (2009) 188–196.
[34] I. Usta, Y. M. Kantar, Mean-variance-skewness-entropy measures: a multi-objective approach for portfolio selection, Entropy 13 (1) (2011) 117–133.
[35] R. Zhou, X. Wang, X. Dong, Z. Zong, Portfolio selection model with the measures of information entropy-incremental entropy-skewness, Advances in Information Sciences and Service Sciences 5 (8) (2013) 833.
[36] W.-G. Zhang, Y.-J. Liu, W.-J. Xu, A possibilistic mean-semivariance-entropy model for multi-period portfolio selection with transaction costs, European Journal of Operational Research 222 (2) (2012) 341–349.
[37] A. K. Bera, S. Y. Park, Optimal portfolio diversification using the maximum entropy principle, Econometric Reviews 27 (4-6) (2008) 484–512.
[38] V. DeMiguel, L. Garlappi, R. Uppal, Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy?, The Review of Financial Studies 22 (5) (2009) 1915–1953.
[39] W. Rödder, I. R. Gartner, S. Rudolph, An entropy-driven expert system shell applied to portfolio selection, Expert Systems with Applications 37 (12) (2010) 7509–7520.
[40] S. Chandrinos, N. D. Lagaros, Construction of currency portfolios by means of an optimized investment strategy, Operations Research Perspectives.
[41] A. Golan, et al., Information and entropy econometricsa review and synthesis, Foundations and trends in econometrics 2 (1–2) (2008) 1–145.
[42] N. Gospodinov, E. Maasoumi, et al., General aggregation of misspecified asset pricing models, Tech. rep., Federal Reserve Bank of Atlanta (2017).
[43] A. Carbone, Information measure for long-range correlated sequences: the case of the 24 human chromosomes, Scientific Reports 3 (2013) 2721.
[44] A. Carbone, H. E. Stanley, Scaling properties and entropy of long-range correlated time series, Physica A: Statistical Mechanics and its Applications 384 (1) (2007) 21–24.

[45] A. Carbone, G. Castelli, H. E. Stanley, Analysis of clusters formed by the moving average of a long-range correlated time series, Phys. Rev. E 69 (2004) 026105.

[46] P. Grassberger, I. Procaccia, Characterization of strange attractors, Physical Review Letters 50 (5) (1983) 346.

[47] S. M. Pincus, Approximate entropy as a measure of system complexity, Proceedings of the National Academy of Sciences 88 (6) (1991) 2297–2301.

[48] C. Bandt, B. Pompe, Permutation entropy: a natural complexity measure for time series, Physical Review Letters 88 (17) (2002) 174102.

[49] O. Rosso, H. Larrondo, M. Martin, A. Plastino, M. Fuentes, Distinguishing noise from chaos, Physical Review Letters 99 (15) (2007) 154102.

[50] J. P. Crutchfield, Between order and chaos, Nature Physics 8 (1) (2012) 17–24.

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

[52] N. Rubido, C. Grebogi, M. S. Baptista, Entropy-based generating markov partitions for complex systems, Chaos: An Interdisciplinary Journal of Nonlinear Science 28 (3) (2018) 033611.

[53] C. S. Daw, C. E. A. Finney, E. R. Tracy, A review of symbolic analysis of experimental data, Review of Scientific Instruments 74 (2) (2003) 915–930.

[54] C. E. Shannon, A mathematical theory of communication, part i, part ii, Bell Syst. Tech. J. 27 (1948) 623–656.

[55] H. Herzel, A. Schmitt, W. Ebeling, Finite sample effects in sequence analysis, Chaos, Solitons & Fractals 4 (1) (1994) 97–113.

[56] P. Grassberger, Finite sample corrections to entropy and dimension estimates, Physics Letters A 128 (6-7) (1988) 369–373.

[57] S. Stammler, S. Katzenbeisser, K. Hamacher, Correcting finite sampling issues in entropy l-diversity, in: International Conference on Privacy in Statistical Databases, Springer, 2016, pp. 135–146.

[58] A. Levina, V. Priesemann, Subsampling scaling, Nature Communications 8 (2017) 15140.

[59] L. Ponta, A. Carbone, S. Cincotti, Detrending moving average algorithm: quantifying heterogeneity in financial data, in: Computer Software and Applications Conference (COMPSAC), 2017 IEEE 41st Annual, Vol. 2, IEEE, 2017, pp. 395–400.

[60] https://it.mathworks.com/help/finance/using-the-sharpe-ratio.html.

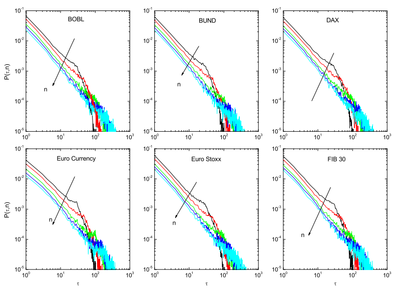

Figure 1
The sketch illustrates the partition obtained by the intersection of the time series $\{x_t\}$ and the moving average $\{\widetilde{x}_{t,n}\}$. Top panel is for moving average window $n = 5$min, bottom panel is for moving average window $n = 10$min. The clusters j, $j+1$, $j+2$, $j+3$ are shown.

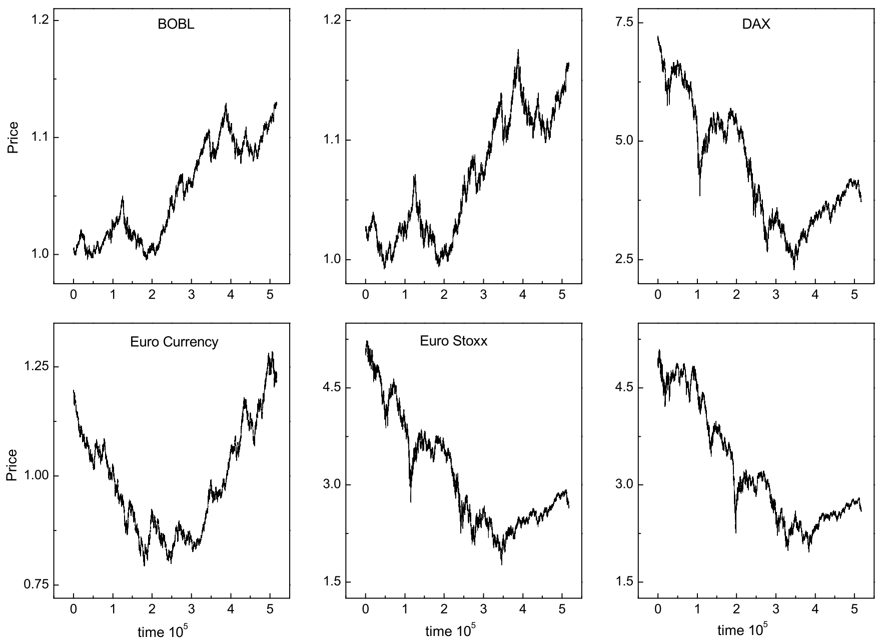

Figure 2
Prices series for the BOBL, BUND, DAX, EuroCurrency, EuroStoxx and FIB30 indexes. The tick-by-tick data series are sampled every minute from 4 January 1999 to 24 March 2004. The total length of the series is 517041.

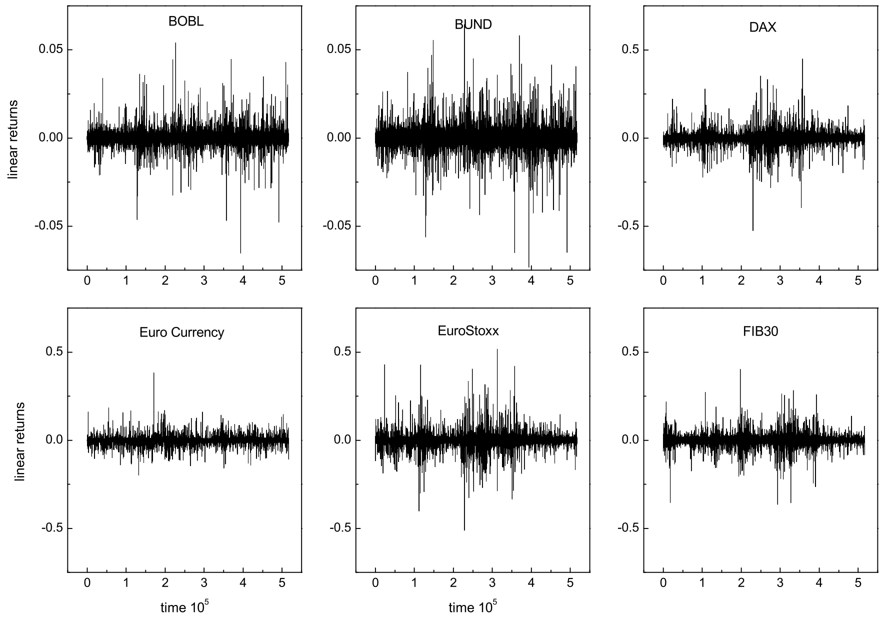

Figure 3
Linear return for the BOBL, BUND, DAX, Euro Currency, Euro Stoxx and FIB30 indexes. The tick-by-tick data series are sampled every minute from 4 January 1999 to 24 March 2004. The total length of the series is 517041.

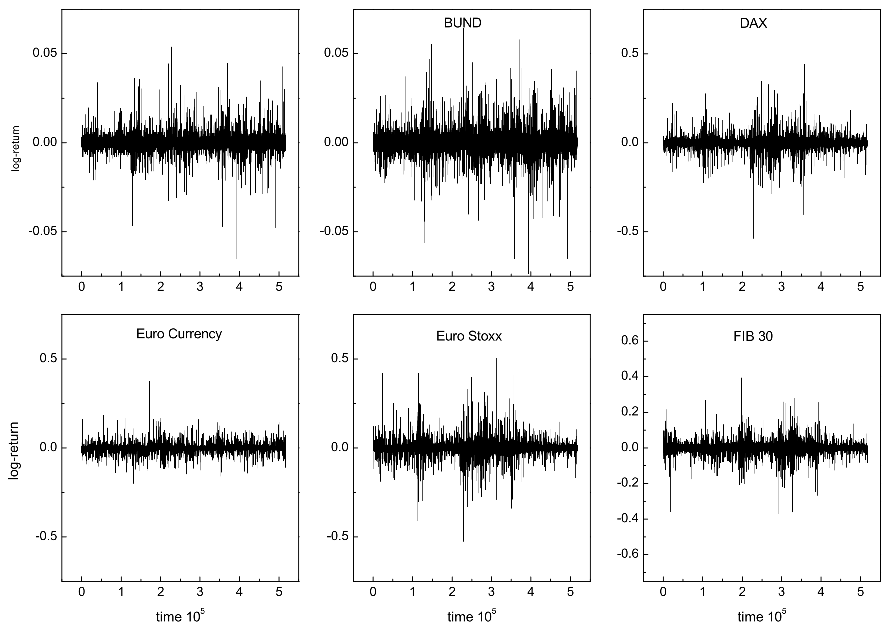

Figure 4
Logarithmic return for the BOBL, BUND, DAX, Euro Currency, Euro Stoxx and FIB30 indexes. The tick-by-tick data series are sampled every minute from 4 January 1999 to 24 March 2004. The total length of the series is 517041.

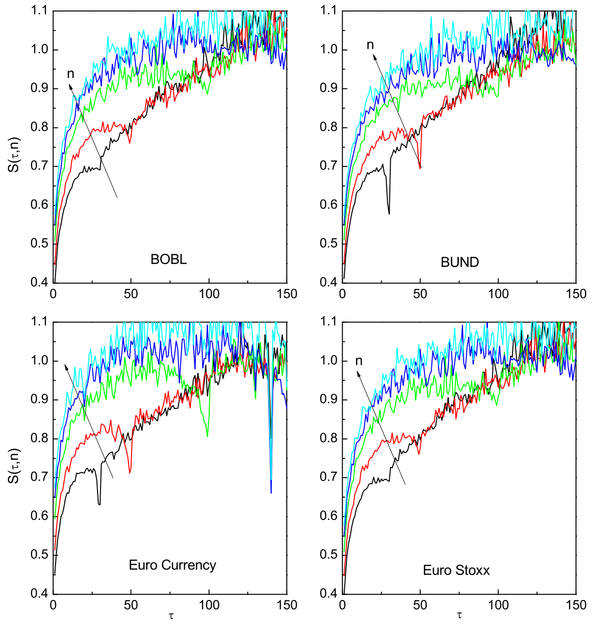

Figure 5
Probability distribution function $P(\tau, n)$ for the prices series of the BOBL, BUND, DAX, Euro Currency, Euro Stoxx and FIB30 markets. The different plots refer to different values of the moving average window n (namely $n = 30$ min, $n = 50$ min, $n = 100$ min, $n = 150$ min and $n = 200$ min).

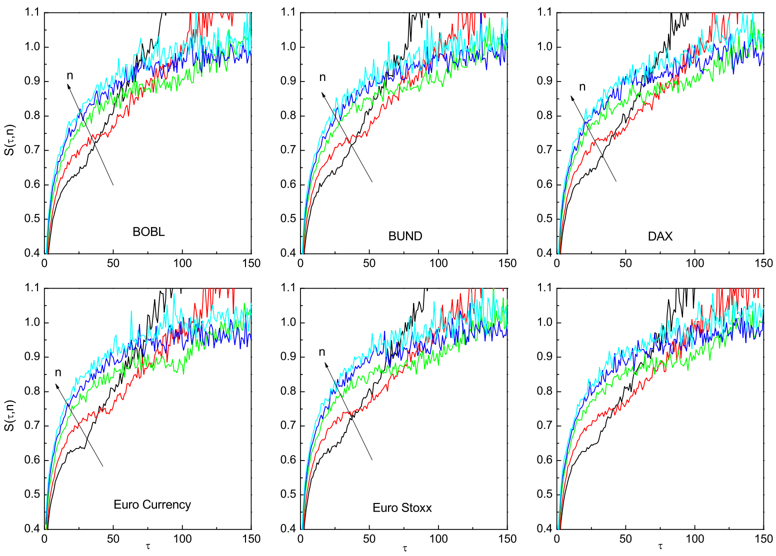

Figure 6
Entropy $S(\tau, n)$ for the time series of the prices of the BOBL, BUND, DAX, Euro Currency, Euro Stoxx and FIB30 markets. The different plots refer to different values of the moving average window n (namely n = 30min, n = 50min, n = 100min, n = 150min and n = 200min).

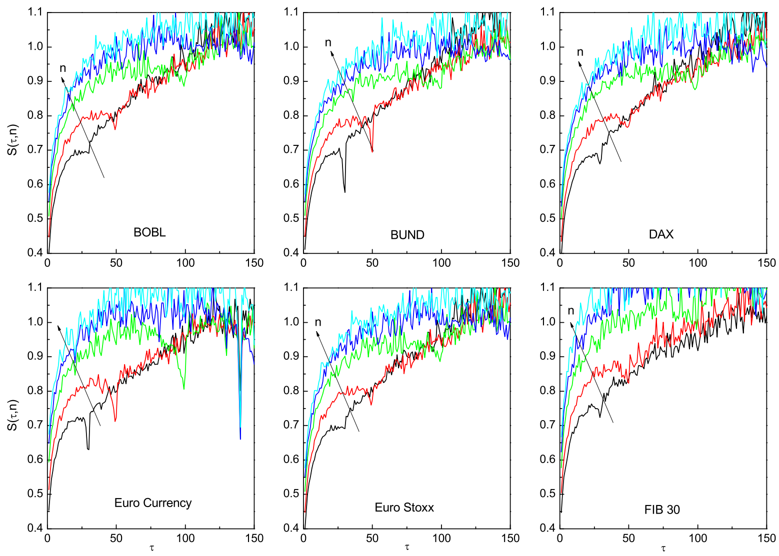

Figure 7
Entropy for the volatility series. The function $S(\tau, n)$ is shown for the volatilities of the log return time series of the BOBL, BUND, DAX, Euro Currency, Euro Stoxx and FIB30 markets. The volatility window $T = 660$ min for all the six graphs. The different plots refer to different values of the moving average window n (namely $n = 30$ min, $n = 50$ min, $n = 100$ min, $n = 150$ min and $n = 200$ min).

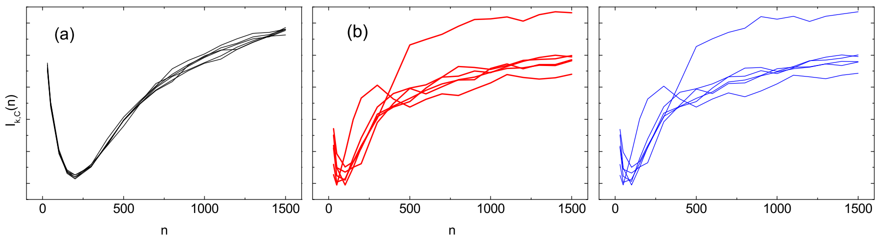

Figure 8
Plot of the function $I_{k,C}(n)$ against n calculated according to Eq. (14) for the prices (a), the volatilities with linear return (b) and with log return (c) for the six indexes. Consistently with the results shown in Fig. (6) the plots exhibit an increasing behaviour with an inversion at small values of n. A strong variability is observed for the volatilities (graphs (b) and (c)) as opposed to the prices series (a).

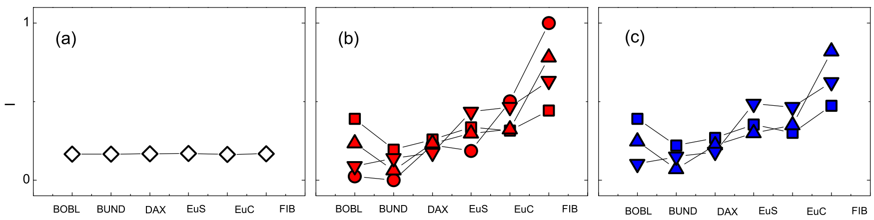

Figure 9
Market Heterogeneity Index $\overline{I}_{k,C}$ calculated according to Eq. (16) for the the six indexes. The volatility window is $T = 1320$ min (circles), $T = 1980$ min (squares), $T = 2640$ min (up triangles) and $T = 3300$ min (down triangles). A strong variability of the $\overline{I}_{k,C}$ is observed for the volatility series (b) and (c), as opposed to the price series exhibiting a constant value of the index (a). The plotted data have been obtained by rescaling $I_{k,C}$ between 0 and 1.

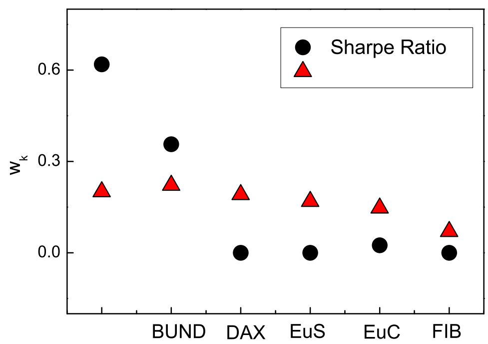

Figure 10
Portfolio weights w_k obtained by using respectively the Sharpe ratio Eq. (12) (circles) and the Market Heterogeneity Index (up triangles) defined by Eq. (17) for the six indexes. One can notice that the triangles are much more diversified that the circles. Furthermore the values of cluster entropy weights are much closer to the maximal entropy weights corresponding to equally distributed assets with $w_k = 1/M = 1/6$.