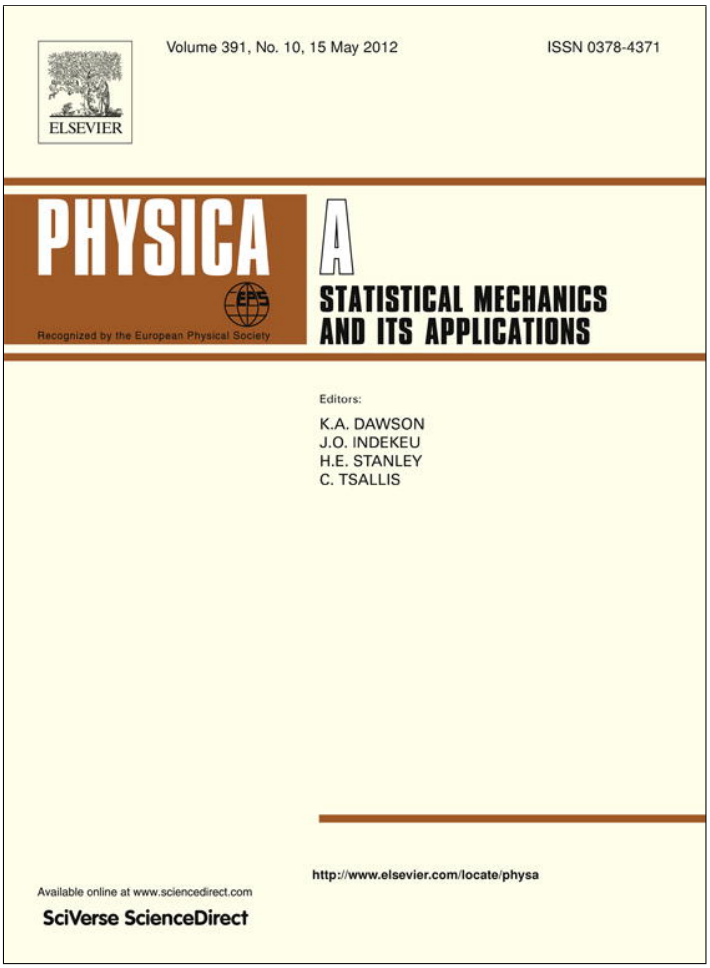

Physica A 391 (2012) 2971–2989

Contents lists available at SciVerse ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Rényi's information transfer between financial time series

Petr Jizba [a,b,*], Hagen Kleinert [a,c], Mohammad Shefaat [d]

[a] ITP, Freie Universität Berlin, Arnimallee 14 D-14195 Berlin, Germany
[b] FNSPE, Czech Technical University in Prague, Břehová 7, 115 19 Praha 1, Czech Republic
[c] ICRANeT, Piazzale della Republica 1, 10 -65122, Pescara, Italy
[d] Quirin Bank AG, Kurfüstendamm 119, 10711 Berlin, Germany

ARTICLE INFO

Article history:
Received 8 July 2011
Received in revised form 9 December 2011
Available online 10 January 2012

Keywords:
Econophysics
Rényi entropy
Information transfer
Financial time series

ABSTRACT

In this paper, we quantify the statistical coherence between financial time series by means of the Rényi entropy. With the help of Campbell's coding theorem, we show that the Rényi entropy selectively emphasizes only certain sectors of the underlying empirical distribution while strongly suppressing others. This accentuation is controlled with Rényi's parameter q. To tackle the issue of the information flow between time series, we formulate the concept of Rényi's transfer entropy as a measure of information that is transferred only between certain parts of underlying distributions. This is particularly pertinent in financial time series, where the knowledge of marginal events such as spikes or sudden jumps is of a crucial importance. We apply the Rényian information flow to stock market time series from 11 world stock indices as sampled at a daily rate in the time period 02.01.1990–31.12.2009. Corresponding *heat maps* and *net information flows* are represented graphically. A detailed discussion of the transfer entropy between the DAX and S&P500 indices based on minute tick data gathered in the period 02.04.2008–11.09.2009 is also provided. Our analysis shows that the bivariate information flow between world markets is strongly asymmetric with a distinct information surplus flowing from the Asia–Pacific region to both European and US markets. An important yet less dramatic excess of information also flows from Europe to the US. This is particularly clearly seen from a careful analysis of Rényi information flow between the DAX and S&P500 indices.

1. Introduction

The evolution of many complex systems in natural, economical, and social sciences is usually presented in the form of time series. In order to analyze time series, several statistical measures have been introduced in the literature. These include such concepts as probability distributions [1,2], autocorrelations [2], multi-fractals [3], complexity [4,5], or entropy densities [5]. Recently, it has been pointed out that the *transfer entropy* (TE) is a very useful instrument in quantifying statistical coherence between time-evolving statistical systems [6–8]. In particular, in Schreiber's paper [6], it was demonstrated that TE is especially expedient when the global properties of time series are analyzed. Prominent applications are in multivariate analysis of time series, including for example, the study of multichannel physiological data or bivariate analysis of historical stock exchange indices. Methods based on TE have substantial computational advantages which are particularly important when analyzing a large amount of data. In all past works, including [8–10], the emphasis has been on various generalizations of transfer entropies that were firmly rooted in the framework of Shannon's information theory. These so-called Shannonian transfer entropies are, indeed, natural candidates due to their ability to quantify in a non-parametric and

* Corresponding author at: ITP, Freie Universität Berlin, Arnimallee 14 D-14195 Berlin, Germany. Tel.: +49 420224358295.
E-mail addresses: jizba@physik.fu-berlin.de, p.jizba@fjfi.cvut.cz (P. Jizba), kleinert@physik.fu-berlin.de (H. Kleinert), mohammad.shefaat@quirinbank.de (M. Shefaat).

0378-4371/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2011.12.064

explicitly non-symmetric way the flow of information between two time series. Financial market time series are an ideal testing ground for various TE concepts, because of the immense amount of electronically recorded financial data.

Recently, economics has become an active research area for physicists. They have investigated stock markets using statistical physics methods, such as percolation theory, multifractals, spin-glass models, information theory, complex networks, and path integrals. In this context, the name econophysics has been coined to denote this new hybrid field on the border between statistical physics and (quantitative) finance. In the framework of econophysics it has became increasingly evident that the market interactions are highly nonlinear, unstable, and long ranged. It has also became apparent that all agents (e.g., companies) involved in a given stock market exhibit interconnectedness and correlations which represent important internal forces of the market. Typically one uses correlation functions to study the internal cross-correlations between various market activities. The correlation functions, however, have at least two limitations. First, they measure only linear relations, although it is clear that linear models do not faithfully reflect real market interactions. Second, all they determine is whether two time series (e.g., two stock-index series) have correlated movement. However, they do not indicate which series affects which, or, in other words, they do not provide any directional information about cause and effect. Some authors use such concepts as time-delayed correlation or time-delayed mutual information in order to construct asymmetric "correlation" matrices with inherent directionality. This procedure is in many respects *ad hoc*, as it does not provide any natural measure (or quantifier) of the information flow between involved series.

In this paper, we study multivariate properties of stock-index time series with the help of an econophysics paradigm. In order to quantify the information flow between two or more stock indices, we generalize Schreibers' Shannonian transfer entropy to Rényi's information setting. With this we demonstrate that the corresponding new transfer entropy provides more detailed information concerning the excess (or lack) of information in various parts of the underlying distribution resulting from updating the distribution on the condition that a second time series is known. This is particularly relevant in the context of financial time series, where the knowledge of tail-part (or marginal) events such as spikes or sudden jumps bears direct implications, e.g., in various risk-reducing formulas in portfolio theory.

The paper is organized as follows. In Section 2, we provide some information-theoretic background on the Shannon entropy (SE) and the Rényi entropy (RE). In particular, we identify the *conditional* Rényi entropy with the information measure introduced in Ref. [11]. Apart from satisfying the chain rule (i.e., rule of additivity of information), the latter has many desirable properties that are to be expected from a conditional information measure. Another key concept, the *mutual* Rényi entropy, is then introduced in a close analogy with Shannon's case. The ensuing properties are also discussed. The Shannonian *transfer* entropy of Schreiber is briefly reviewed in Section 3. There we also comment on the effective transfer entropy of Marschinski et al. The core quantity of this work, the Rényian *transfer* entropy (RTE), is motivated and derived in Section 4. In contrast to the Shannonian case, the Rényian transfer entropy is generally not positive semi-definite. This is because the RE nonlinearly emphasizes different parts of the probability density functions (PDFs) involved. With the help of Campbell's coding theorem, we show that the RTE rates a gain/loss in risk involved in a next-time-step behavior in a given stochastic process, say X, resulting from learning new information, namely the historical behavior of another (generally cross-correlated) process, say Y. In this view, the RTE can serve as a convenient *rating* factor of a *riskiness* in interconnected markets. We also show that the Rényian transfer entropy allows one to amend spurious effects caused by the finite size of a real data set which in Shannon's context must, otherwise, be solved by means of the surrogate data technique and ensuing effective transfer entropy. In Section 5, we demonstrate the usefulness and formal consistency of the RTE by analyzing cross-correlations in various international stock markets. On a qualitative level, we use 183,308 simultaneously recorded data points of the eleven stock exchange indices, sampled at a daily (end of trading day) rate to construct the *heat maps* and *net flows* for both Shannon's and Rényi's information flows. On a quantitative level, we explicitly discuss time series from the DAX and S&P500 market indices gathered on a minute-tick basis in the period from December 1990 to November 2009 in the German stock exchange market (Deutsche Börse). Presented calculations of Rényi and Shannon transfer entropies are based on symbolic coding computation with the open source software R. Our numerical results imply that the RTE values among world markets are typically very asymmetric. For instance, we show that there is a strong surplus of information flow from the Asia–Pacific region to both Europe and the US. A surplus of information flow can be also observed to exist from Europe to the US. In this last case, the substantial volume of transferred information comes from the tail part (i.e., the risky part) of the underlying asset distributions. So, despite the fact that the US contributes more than half of the world's trading volume, this is not so with information flow.

Further salient issues, such as the dependence of RTE on Rényi's q parameter or on the data block length, are also investigated numerically. In this context we find that the cross-correlation between the DAX and S&P500 indices has a long-time memory which is around 200–300 min. This should be contrasted with the typical memory of stock returns, which are of the order of seconds or maximally a few minutes. Various remarks and generalizations are proposed in the concluding section, Section 6. For the reader's convenience, we give in Appendix A a brief glossary of market indices used in the main text, and in Appendix B we tabulate explicit values of the effective transfer entropies used in the construction of heat maps and net information flows.

2. Information-theoretic entropies of Shannon and Rényi

In order to express numerically an amount of information that is shared or transferred between various data sets (e.g., two or more random processes), one commonly resorts to information theory, and especially to the concept of entropy. In

this section, we briefly review some essentials of Shannon's entropy and Rényi's entropy that will be needed in the following sections.

2.1. Shannon's entropy

The entropy concept was originally introduced by Clausius [12] in the framework of thermodynamics. By analyzing a Carnot engine, he was able to identify a new state function which never decreases in isolated systems. The microphysical origin of Clausius' phenomenological entropy was clarified more than 20 years later in works of Boltzmann and (even later) Gibbs, who associated Clausius entropy with the number of allowed microscopic states compatible with a given observed macrostate. The ensuing *Boltzmann–Gibbs entropy* reads

$$H_{\mathrm{BG}}(\mathcal{P}) = -k_B \sum_{x\in X}^{W} p(x)\ln p(x), \tag{1}$$

where k_B is Boltzmann's constant, X is the set of all accessible microstates compatible with whatever macroscopic observable (state variable) one controls, and W denotes the number of microstates.

It should be said that the passage from Boltzmann–Gibbs to Clausius entropy is established only when the conditional extremum $\mathcal{P}_{\mathrm{ex}}$ of H_{BG} subject to the constraints imposed by observed state variables is inserted back into H_{BG}. Only when this *maximal entropy prescription* [13] is utilized does H_{BG} turn out to be a thermodynamic state function and not a mere functional on a probability space.

In information theory, on the other hand, the interest was in an optimal coding of a given source data. By *optimal code* is meant the shortest averaged code from which one can uniquely decode the source data. Optimality of coding was solved by Shannon in his seminal 1948 paper [14]. According to Shannon's *source coding theorem* [14,15], the quantity

$$H(\mathcal{P}) = -\sum_{x\in X}^{W} p(x)\log_2 p(x) \tag{2}$$

corresponds to the averaged number of bits needed to optimally encode (or "zip") the source dataset X with the source probability distribution $\mathcal{P}(X)$. On a quantitative level, (2) represents (in bits) the minimal number of binary (yes/no) questions that brings us from our present state of knowledge about the system X to the one of certainty [14,16,17]. It should be stressed that, in Shannon's formulation, X represents a discrete set (e.g., processes with discrete time), and this will also be the case here. Apart from the foregoing *operational* definitions, Eq. (2) also has several axiomatic underpinnings. Axiomatic approaches have been advanced by Shannon [14,15], Khinchin [18], Feinstein [19] and others [20]. The quantity (2) has become known as the Shannon entropy.

There is an intimate connection between the Boltzmann–Gibbs entropy and the Shannon entropy. In fact, thermodynamics can be viewed as a specific application of Shannon's information theory: the thermodynamic entropy may be interpreted (when rescaled to "bit" units) as the amount of Shannon information needed to define the detailed microscopic state of the system, which remains "uncommunicated" by a description that is solely in terms of thermodynamic state variables [21–23].

Among important properties of the SE is its concavity in $\mathcal{P}$; i.e., for any pair of distributions $\mathcal{P}$ and $\mathcal{Q}$, and a real number $0 \le \lambda \le 1$, it holds that

$$H(\lambda\mathcal{P} + (1-\lambda)\mathcal{Q}) \ge \lambda H(\mathcal{P}) + (1-\lambda)H(\mathcal{Q}). \tag{3}$$

Eq. (3) follows from Jensen's inequality and the convexity of $x\log x$ for $x > 0$. Concavity is an important concept since it ensures that any maximizer found by the methods of the differential calculus yields an absolute maximum rather than a relative maximum or minimum or saddle point. At the same time it is just a sufficient (i.e., not necessary) condition guarantying a unique maximizer. It is often customary to denote the SE of the source X as $H(X)$ rather than $H(\mathcal{P})$. Note that the SE is generally not convex in X!

It should be stressed that the entropy (2) really represents self-information: the information yielded by a random process about itself. A step further from self-information offers the *joint entropy* of two random variables X and Y, which is defined as

$$H(X\cap Y) = -\sum_{x\in X, y\in Y} p(x,y)\log_2 p(x,y), \tag{4}$$

and which represents the amount of information gained by observing jointly two (generally dependent or correlated) statistical events.

A further concept that will be needed here is the *conditional entropy* of X given Y, which can be motivated as follows. Let us have two statistical events X and Y and let event Y have a sharp value y. Then the gain of information obtained by observing X is

$$H(X|Y=y) = -\sum_{x\in X} p(x|y)\log_2 p(x|y). \tag{5}$$

Here, the conditional probability $p(x|y) = p(x, y)/p(y)$. For general random Y one defines the conditional entropy as the averaged SE yielded by X under the assumption that the value of Y is known; i.e.,

$$H(X|Y) = \sum_{y\in Y} p(y)H(X|Y = y) = -\sum_{x\in X, y\in Y} p(x, y)\log_2 p(x|y). \tag{6}$$

From (6), in particular, it follows that

$$H(X \cap Y) = H(Y) + H(X|Y) = H(X) + H(Y|X). \tag{7}$$

Identity (7) is known as the additivity (or chain) rule for the Shannon entropy. In statistical thermodynamics, this rule allows one to explain, for example, the Gibbs paradox. Applying Eq. (7) iteratively, we obtain

$$\begin{aligned} H(X_1 \cap X_2 \cap \cdots \cap X_n) &= H(X_1) + H(X_2|X_1) + H(X_3|X_1 \cap X_2) + \cdots \\ &= \sum_{i}^{n} H(X_i|X_1 \cap X_2 \cap \cdots \cap X_{i-1}). \end{aligned} \tag{8}$$

Another relevant quantity that will be needed is the *mutual information* between X and Y. This is defined as

$$I(X; Y) = \sum_{x\in X, y\in Y} p(x, y)\log_2 \frac{p(x, y)}{p(x)q(y)}, \tag{9}$$

and can be equivalently written as

$$I(X; Y) = H(X) - H(X|Y) = H(Y) - H(Y|X). \tag{10}$$

This shows that the mutual information measures the average reduction in uncertainty (i.e., gain in information) about X resulting from observation of Y. Of course, the amount of information contained in X about itself is just the Shannon entropy:

$$I(X; X) = H(X). \tag{11}$$

Notice also that from Eq. (9) it follows that $I(X; Y) = I(Y; X)$, and so X provides the same amount of information on Y as Y does on X. For this reason, mutual information is not a useful measure to quantify a flow of information. In fact, the flow of information should be by its very definition directional.

In the following, we will also find useful the concept of *conditional mutual entropy* between X and Y given Z, which is defined as

$$\begin{aligned} I(X; Y|Z) &= H(X|Z) - H(X|Y \cap Z), \\ &= I(X; Y \cap Z) - I(X; Y). \end{aligned} \tag{12}$$

The latter quantifies the averaged mutual information between X and Y provided that Z is known. Applying (12) and (10) iteratively, we may write

$$\begin{aligned} I(X; Y_1 \cap \cdots \cap Y_n|Z_1 \cap \cdots \cap Z_m) &= H(X|Z_1 \cap \cdots \cap Z_m) - H(X|Y_1 \cap \cdots \cap Y_n \cap Z_1 \cap \cdots \cap Z_m) \\ &= I(X; Y_1 \cap \cdots \cap Y_n \cap Z_1 \cap \cdots \cap Z_m) - I(X; Z_1 \cap \cdots \cap Z_m). \end{aligned} \tag{13}$$

For further details on the basic concepts of Shannon's information theory, we refer the reader to classical books, e.g., [17] and, more recently, Csiszár and Shields [24].

2.2. Rényi's entropy

Rényi introduced in Refs. [25,26] a one-parameter family of information measures currently known as *Rényi entropies* [11,25]. In practice, however, only a singular name – Rényi's entropy – is used. The Rényi entropy of order q ($q > 0$) of a distribution $\mathcal{P}$ on a finite set X is defined as

$$S_q^{(R)}(\mathcal{P}) = \frac{1}{1-q}\log_2 \sum_{x\in X} p^q(x). \tag{14}$$

For RE (14), one can also formulate a source coding theorem. While in the Shannon case the cost of a codeword is a linear function of the length, so the optimal code has a minimal cost out of all codes, in the Rényi case the cost of a codeword is an exponential function of its length [27–29]. This is, in a nutshell, the essence of the so-called Campbell coding theorem (CCT). According to this, the RE corresponds to the averaged number of bits needed to optimally encode the discrete source X with the probability $\mathcal{P}(X)$, provided that the codeword lengths are exponentially weighted [30]. From the form (14), one can easily see that for $q > 1$ the RE depends more on the probabilities of the more probable values and less on those of the improbable ones. This dependence is more pronounced for higher q. On the other hand, for $0 < q < 1$, marginal events

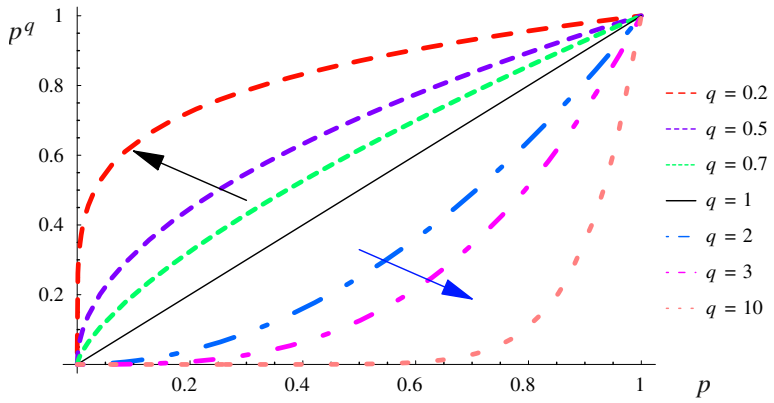

Fig. 1. The function p^q for event probability p and varying Rényi parameter q. Arrows indicate decreasing values of q for $0 < q < 1$ (dark arrow) or increasing values of q for $q > 1$ (lighter arrow).

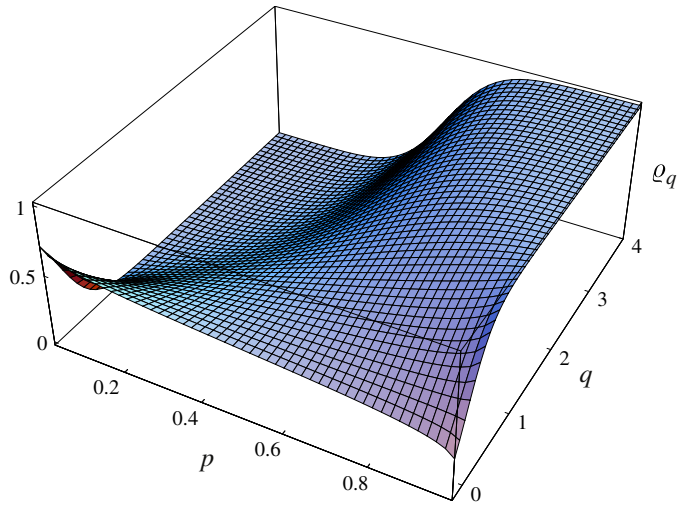

Fig. 2. A plot of the escort distribution for two-dimensional $\mathcal{P}$: $\varrho_q = p^q/(p^q + (1-p)^q)$.

are accentuated with decreasing q. In this context, we should also point out that Campbell's coding theorem for the RE is equivalent to Shannon's coding theorem for the SE provided one uses instead of $p(x)$ the *escort distribution* [29]:

$$\varrho_q(x) \equiv \frac{p^q(x)}{\sum_{x\in X} p^q(x)}. \tag{15}$$

The PDF $\varrho_q(x)$ was first introduced by Rényi [26] and in the physical context by Beck et al. and others (see, e.g., Refs. [31,32]). Note (see Fig. 1) that for $q > 1$ the escort distribution emphasizes the more probable events and de-emphasizes the more improbable ones. This trend is more pronounced for higher values of q. For $0 < q < 1$, the escort distribution accentuates more improbable (i.e., marginal or rare) events. This dependence is more pronounced for decreasing q. This fact is clearly seen in Fig. 2. So, by choosing different q, we can "scan" or "probe" different parts of the PDFs involved.

It should be stressed that, apart from the CCT, the RE has yet further operational definitions, e.g., in the theory of guessing [33], in the buffer overflow problem [34], or in the theory of error block coding [35]. The RE is also underpinned with various axiomatics [25,26,36]. In particular, it satisfies identical Khinchin axioms [18] as Shannon's entropy save for the additivity axiom (chain rule) [11,37,38]:

$$S_q^{(R)}(X \cap Y) = S_q^{(R)}(Y) + S_q^{(R)}(X|Y), \tag{16}$$

where the conditional entropy $S_q^{(R)}(X|Y)$ is defined with the help of the escort distribution (15) (see, e.g., Refs. [11,31,39]). For $q \to 1$, the RE reduces to the SE:

$$S_1^{(R)} = \lim_{q\to 1} S_q^{(R)} = H, \tag{17}$$

as one can easily verify with l'Hospital's rule.

We define the *joint Rényi entropy* (or the *joint entropy* of order q) for two random variables X and Y in a natural way as

$$S_q^{(R)}(X \cap Y) = \frac{1}{1-q} \log_2 \sum_{x \in X} p^q(x, y). \tag{18}$$

The *conditional entropy* of order q of X given Y is similarly as in the Shannon case defined as the averaged Rényi entropy yielded by X under the assumption that the value of Y is known. As shown in Refs. [11,37,40], this has the form

$$\begin{aligned} S_q^{(R)}(X|Y) &= \frac{1}{1-q} \log_2 \frac{\sum_{x \in X, y \in Y} p^q(x|y) q^q(y)}{\sum_{y \in Y} q^q(y)} \\ &= \frac{1}{1-q} \log_2 \frac{\sum_{x \in X, y \in Y} p^q(x, y)}{\sum_{y \in Y} q^q(y)}. \end{aligned} \tag{19}$$

In this context, it should be mentioned that several alternative definitions of the conditional RE exist (see, e.g., Refs. [26,35,41]), but the formulation (19) differs from other versions in a few important ways that will be shown to be desirable in the following considerations. The conditional entropy defined in (19) has the following important properties [11,40]:

- $0 \leq S_q^{(R)}(X|Y) \leq \log_2 n$, where n is the number of elements in X,
- $S_q^{(R)}(X|Y) = 0$ only when Y uniquely determines X (i.e., no gain in information),
- $\lim_{q \to 1} S_q^{(R)}(X|Y) = H(X|Y)$,
- when X and Y are independent, then $S_q^{(R)}(X|Y) = S_q^{(R)}(X)$.

Unlike in the Shannon case, one cannot, however, deduce that the equality $S_q^{(R)}(X|Y) = S_q^{(R)}(X)$ implies independence between events X and Y. Also the inequality $S_q^{(R)}(X|Y) \leq S_q^{(R)}(X)$ (i.e., an extra knowledge about Y lessens our ignorance about X) does not hold here in general [11,26]. The latter two properties may seem to be a serious flaw. We will now argue that this is not the case and, in fact, it is even desirable.

First, in order to understand why $S_q^{(R)}(X|Y) = S_q^{(R)}(X)$ does not imply independence between X and Y, we define the information-distribution function

$$\mathcal{F}_{\mathcal{P}}(x) = \sum_{-\log_2 p(z) < x} p(z), \tag{20}$$

which represents the total probability caused by events with information content $H(z) = -\log_2 p(z) < x$. With this we have

$$2^{(1-q)x} \mathrm{d}\mathcal{F}_{\mathcal{P}}(x) = \sum_{x \leq H(z) < x + \mathrm{d}x} 2^{(1-q)H(z)} p(z) = \sum_{x \leq H(z) < x + \mathrm{d}x} p^q(z), \tag{21}$$

and thus

$$S_q^{(R)}(X) = \frac{1}{1-q} \log_2 \left(\int_0^\infty 2^{(1-q)x} \mathrm{d}\mathcal{F}_{\mathcal{P}}(x) \right). \tag{22}$$

Taking the inverse Laplace transform with the help of the so-called *Post inversion formula* [42], we obtain

$$\mathcal{F}_{\mathcal{P}}(x) = \lim_{k \to \infty} \frac{(-1)^k}{k!} \left(\frac{k}{x \ln 2} \right)^{k+1} \frac{\partial^k}{\partial q^k} \left[\frac{2^{(1-q) S_q^{(R)}(X)}}{(q-1)} \right] \Bigg|_{q = k/(x \ln 2) + 1}. \tag{23}$$

An analogous relation also holds for $\mathcal{F}_{\mathcal{P}|\mathcal{Q}}(x)$ and associated $S_q^{(R)}(X|Y)$. As a result, we see that when working with $S_q^{(R)}$ of different orders we receive much more information on the underlying distribution than when we restrict our investigation to only one q (e.g., to only the Shannon entropy). In addition, Eq. (23) indicates that we need all $q > 1$ (or equivalently all $0 < q < 1$; see Ref. [43]) in order to uniquely identify the underlying PDF.

In view of Eq. (23), we see that the equality between $S_q^{(R)}(X|Y)$ and $S_q^{(R)}(X)$ at some neighborhood of q merely implies that $\mathcal{F}_{\mathcal{P}|\mathcal{Q}}(x) = \mathcal{F}_{\mathcal{P}}(x)$ for some x. This naturally does not ensure independence between X and Y. We need the equality $S_q^{(R)}(X|Y) = S_q^{(R)}(X)$ for all $q > 1$ (or for all $0 < q < 1$) in order to secure that $\mathcal{F}_{\mathcal{P}|\mathcal{Q}}(x) = \mathcal{F}_{\mathcal{P}}(x)$ holds for all x, which would in turn guarantee that $\mathcal{P}(X) = \mathcal{P}(X|Y)$. Therefore, all REs with $q > 1$ (or all with $0 < q < 1$) are generally required to deduce from $S_q^{(R)}(X|Y) = S_q^{(R)}(X)$ an independence between X and Y.

In order to understand the meaning of the inequality $S_q^{(R)}(X|Y) \leq S_q^{(R)}(X)$, we first introduce the concept of mutual information. The *mutual information of order q* between X and Y can be defined as (see Eq. (10))

$$\begin{aligned} I_q^{(R)}(X;Y) &= S_q^{(R)}(X) - S_q^{(R)}(X|Y) \\ &= S_q^{(R)}(X) + S_q^{(R)}(Y) - S_q^{(R)}(X \cap Y), \end{aligned} \tag{24}$$

which explicitly reads

$$\begin{aligned} I_q^{(R)}(X;Y) &= \frac{1}{1-q} \log_2 \frac{\sum\limits_{x\in X, y\in Y} q^q(y) p^q(x)}{\sum\limits_{x\in X, y\in Y} p^q(x,y)} \\ &= \frac{1}{1-q} \log_2 \frac{\sum\limits_{x\in X, y\in Y} q^q(y) p^q(x)}{\sum\limits_{x\in X, y\in Y} q^q(y) p^q(x|y)}. \end{aligned} \tag{25}$$

Note that we have again the symmetry relation $I_q^{(R)}(X;Y) = I_q^{(R)}(Y;X)$ as well as the consistency condition $I_q^{(R)}(X;X) = S_q^{(R)}(X)$. So, similarly as in the Shannon case, Rényi's mutual information formally quantifies the average reduction in uncertainty (i.e., gain in information) about X that results from learning the value of Y, or vice versa.

From Eq. (24), we see that the inequality in question, i.e., $S_q^{(R)}(X|Y) \leq S_q^{(R)}(X)$, implies that $I_q^{(R)}(Y;X) \geq 0$. According to (25), this can be violated only when

$$\begin{aligned} &\sum_{x\in X} p^q(x) > \sum_{x\in X} \langle \mathcal{P}^q(x|Y)\rangle_q \quad \text{for } q > 1, \\ &\sum_{x\in X} p^q(x) < \sum_{x\in X} \langle \mathcal{P}^q(x|Y)\rangle_q \quad \text{for } 0 < q < 1. \end{aligned} \tag{26}$$

Here, $\langle \ldots \rangle_q$ is an average with respect to the escort distribution $\varrho_q(y)$ (see Eq. (15)).

By taking into account properties of the escort distribution, we can deduce that $I_q^{(R)}(X;Y) < 0$ when a larger probability events of X obtain a lower value by learning Y. As for the marginal events of X, these are indeed enhanced by learning Y, but the enhancement rate is smaller than the suppression rate of large probabilities. For instance, this happens when

$$\mathcal{P}(X) = \left\{1-\epsilon, \frac{\epsilon}{n-1}, \ldots, \frac{\epsilon}{n-1}\right\} \mapsto \mathcal{P}(X|Y) = \left\{\frac{1-\epsilon}{2}, \frac{1-\epsilon}{2}, \frac{\epsilon}{n-2}, \ldots, \frac{\epsilon}{n-2}\right\}, \tag{27}$$

for

$$\frac{1}{1+\log_2\left(\frac{n-1}{n-2}\right)} \leq \epsilon < 1, \quad n > 2. \tag{28}$$

The inequality (28) ensures that $I(Y;X) \geq 0$ holds. The left inequality in (28) saturates when $I(Y;X) = 0$; see also Fig. 3.

For $0 < q < 1$, the situation is analogous. Here, properties of the escort distribution imply that $I_q^{(R)}(Y;X) < 0$ when marginal events of X obtain a higher probability by learning Y. The suppression rate for large (i.e. close-to-peak) probabilities is now smaller than the enhancement rate of marginal events. This happens, for example, for distributions,

$$\mathcal{P}(X) = \left\{\frac{1-\epsilon}{2}, \frac{1-\epsilon}{2}, \frac{\epsilon}{n-2}, \ldots, \frac{\epsilon}{n-2}\right\} \mapsto \mathcal{P}(X|Y) = \left\{1-\epsilon, \frac{\epsilon}{n-1}, \ldots, \frac{\epsilon}{n-1}\right\}, \tag{29}$$

with ϵ again fulfilling the inequality (28). This can be also directly seen from Fig. 3 when we revert the sign of $I_q^{(R)}(Y;X)$. When we set $q = 1$, then both inequalities (26) are simultaneously satisfied, yielding $I(Y;X) = 0$, as it should.

In contrast to the Shannonian case, in which the mutual information quantifies the average reduction in uncertainty resulting from observing/learning further information, in the Rényi case we should use Campbell's coding theorem in order to properly understand the meaning of $I_q^{(R)}(Y;X)$. According to the CCT $S_q^{(R)}(X)$ corresponds to the minimal average cost of a coded message with a nonlinear (exponential) weighting/pricing of codeword lengths. While according to Shannon we never increase ignorance by learning Y (i.e., possible correlations between X and Y can only reduce the entropy), in Rényi's setting, extra knowledge about Y might easily increase the minimal price of coding X because of the nonlinear pricing. Since the CCT penalizes long codewords which in Shannon's coding have low probability, the price of the $X|Y$ code may easily increase, as we have seen in examples (27) and (29).

In the key context of financial time series, the risk valuation of large changes such as spikes or sudden jumps is of a crucial importance, e.g., in various risk-reducing formulas in portfolio theory. The rôle of Campbell's pricing can in these cases be interpreted as a risk-rating method which puts an exponential premium on rare (i.e., risky) asset fluctuations. From this

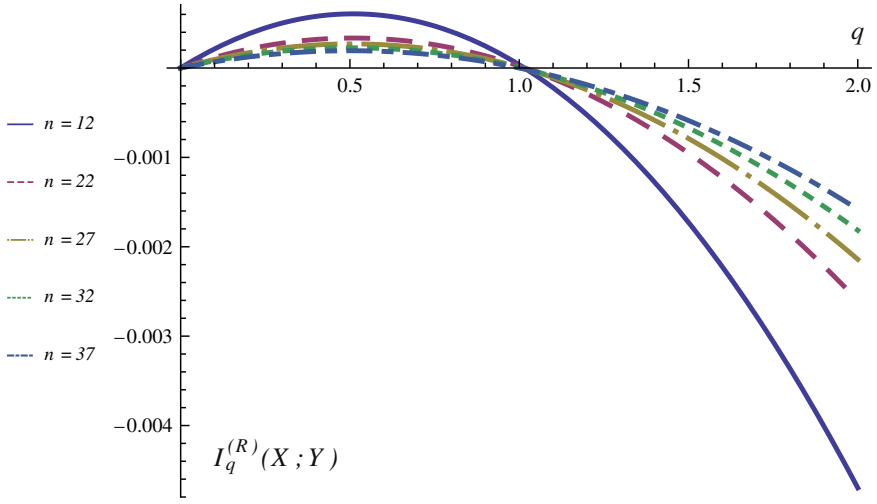

Fig. 3. Example of a typical situation when $I_q^{(R)}(X;Y)$ is negative. Distributions $\mathcal{P}(X)$ and $\mathcal{P}(X|Y)$ are specified in (27), and ϵ is chosen so that correspondingly $I(X;Y)=0$.

point of view, the mutual information $I_q^{(R)}(X;Y)$ represents a *rating factor* which rates a gain/loss in risk in X resulting from learning new information, namely information about Y.

The *conditional mutual information* of order q between X and Y given Z is defined as

$$I_q^{(R)}(X;Y|Z) = S_q^{(R)}(X|Z) - S_q^{(R)}(X|Y\cap Z). \tag{30}$$

Note that because of a validity of the chain rule (16), relations (8) and (13) also hold true for the RE.

To close this section, we shall stress that information entropies are primarily important because there are various coding theorems which endow them with an operational (that is, experimental) meaning, and not because of intuitively pleasing aspects of their definitions. While coding theorems do exist both for the Shannon entropy and the Rényi entropy, there are (as yet) no such theorems for Tsallis', Kaniadakis', Naudts' and other currently popular entropies. The information-theoretic significance of such entropies is thus not obvious. Since the information-theoretic aspect of entropies is of a crucial importance here, we will in the following focus only on the SE and the RE.

3. Fundamentals of Shannonian transfer entropy

3.1. Shannonian transfer entropy

As seen in Section 2.1, the mutual information $I(X;Y)$ quantifies the decrease of uncertainty about X caused by the knowledge of Y. One could be thus tempted to use it as a measure of an informational transfer in general complex systems. A major problem, however, is that Shannon's mutual information contains no inherent directionality, since $I(X;Y)=I(Y;X)$. Some early attempts tried to resolve this complication by artificially introducing the directionality via time-lagged random variables. In this way one may define, for instance, the *time-lagged mutual* (or *directed Kullback–Leibler*) *information* as

$$I(X;Y)_{t,\tau} = \sum p(x_t, y_{t-\tau}) \log_2 \frac{p(x_t, x_{t-\tau})}{p(x_t)q(y_t)}. \tag{31}$$

The later describes the average gain of information when replacing the product probability $\mathcal{P}_t \times \mathcal{Q}_t = \{p(x_t)q(y_t); x_t \in X_t, y_t \in Y_t\}$ by the joint probability $\mathcal{P}_t \cap \mathcal{Q}_{t-\tau} = \{p(x_t, y_{t-\tau}); x_t \in X_t, y_{t-\tau} \in Y_{t-\tau}\}$. So the information gained is due to cross-correlation effect between random variables X_t and Y_t (respectively, $Y_{t-\tau}$). It was, however, pointed out in Ref. [6] that prescriptions such as (31), though directional, also take into account some part of the information that is statically shared between the two random processes X and Y. In other words, these prescriptions do not produce statistical dependences that truly originate only in the stochastic random process Y, but they do include the effects of a common history (such as, for example, in the case of a common external driving force).

For this reason, Schreiber introduced in Ref. [6] the concept of (Shannonian) transfer entropy (STE). The latter, apart from directionality, accounts only for the cross-correlations between statistical time series X and Y whose genuine origin is in the "source" process Y. The essence of the approach is the following. Let us have two time sequences described by stochastic random variables X_t and Y_t. Let us assume further that the time steps (data ticks) are discrete with the size of an elementary time lag τ and with $t_n = t_0 + n\tau$ (t_0 is a reference time).

The transfer entropy $T_{Y\to X}(m,l)$ can then be defined as

$$\begin{aligned} T_{Y\to X}(m,l) &= H(X_{t_{m+1}}|X_{t_1}\cap\cdots\cap X_{t_m}) - H(X_{t_{m+1}}|X_{t_1}\cap\cdots\cap X_{t_m}\cap Y_{t_{m-l+1}}\cap\cdots\cap Y_{t_m}) \\ &= I(X_{t_{m+1}}; X_{t_1}\cap\cdots\cap X_{t_m}\cap Y_{t_{m-l+1}}\cap\cdots\cap Y_{t_m}) - I(X_{t_{m+1}}; X_{t_1}\cap\cdots\cap X_{t_m}). \end{aligned} \tag{32}$$

The last line of (32) indicates that $T_{Y\to X}(m,l)$ represents the following.

$+$ Gain of information about $X_{t_{m+1}}$ caused by the whole history of X and Y up to time t_m
$-$ Gain of information about $X_{t_{m+1}}$ caused by the whole history of X up to time t_m
$=$ Gain of information about $X_{t_{m+1}}$ caused purely by the whole history of Y up to time t_m.

Note that one may equivalently rewrite (32) as the conditional mutual information

$$T_{Y\to X}(m,l) = I(X_{t_{m+1}}; Y_{t_{m-l+1}} \cap \cdots \cap Y_{t_m} | X_{t_1} \cap \cdots \cap X_{t_m}). \tag{33}$$

This shows once more the essence of Schreiber's transfer entropy, namely, that it describes the gain in information about $X_{t_{m+1}}$ caused by the whole history of Y (up to time t_m) under the assumption that the whole history of X (up to time t_m) is known. According to the definition of the conditional mutual information, we can explicitly rewrite Eq. (33) as

$$T_{Y\to X}(m,l) = \sum p(x_{t_1},\ldots,x_{t_{m+1}},y_{t_{m-l+1}},\ldots,y_{t_m}) \log_2 \frac{p(x_{t_{m+1}}|x_{t_1},\ldots,x_{t_m},y_{t_{m-l+1}},\ldots,y_{t_m})}{p(x_{t_{m+1}}|x_{t_1},\ldots,x_{t_m})}, \tag{34}$$

where x_t and y_t represent the discrete states at time t of X and Y, respectively.

In passing, we may observe from the first line of (32) that $T_{Y\to X} \geq 0$ (any extra knowledge in conditional entropy lessens the ignorance). In addition, due to the Shannon–Gibbs inequality (see, e.g., Ref. [23]), $T_{Y\to X} = 0$ only when

$$\frac{p(x_{t_{m+1}}|x_{t_1},\ldots,x_{t_m},y_{t_{m-l+1}},\ldots,y_{t_m})}{p(x_{t_{m+1}}|x_{t_1},\ldots,x_{t_m})} = 1. \tag{35}$$

This, however, means that the history of Y up to time t_m has no influence on the value of $X_{t_{m+1}}$ or, in other words, there is no information flow from Y to X; i.e., the Y and X time series are independent processes. If there is any kind of information flow, then $T_{Y\to X} > 0$. $T_{Y\to X}$ is clearly explicitly non-symmetric (directional) since it measures the degree of dependence of X on Y and not vice versa.

3.2. Effective transfer entropy

The effective transfer entropy (ETE) was originally introduced by Marschinski et al. in Ref. [7], and it was further substantiated in Refs. [9,10,44]. The ETE, in contrast to the STE, accounts for the finite size of a real data set.

In the previous section, we have defined $T_{Y\to X}(m,l)$ with the history indices m and l. In order to view $T_{Y\to X}$ as a genuine transfer entropy, one should really include in (33) the whole history of Y and X up to time t_m (i.e., all historical data that may be responsible for cross-correlations with $X_{t_{m+1}}$). The history is finite only if X or/and Y processes are Markovian. In particular, if X is a Markov process of order $m+1$ and Y is of order l, then $T_{Y\to X}(m,l)$ is a true transfer entropy. Unfortunately, most dynamical systems cannot be mapped to Markovian processes with finite-time memory. For such systems one should take limits $m\to\infty$ and $l\to\infty$. In practice, however, the finite size of any real data set hinders this limiting procedure. In order to avoid unwanted finite-size effects, Marschinski proposed the quantity

$$T^{\text{eff}}_{Y\to X}(m,l) \equiv T_{Y\to X}(m,l) - T_{Y_{\text{schuffled}}\to X}(m,l), \tag{36}$$

where $Y_{\text{schuffled}}$ indicates data shuffling via the *surrogate data* technique [45]. The surrogate data sequence has the same mean, the same variance, the same autocorrelation function, and therefore the same power spectrum as the original sequence, but (nonlinear) phase relations are destroyed. In effect, all the potential correlations between time series X and Y are removed, which means that $T_{Y_{\text{schuffled}}\to X}(m,l)$ should be zero. In practice, this shows itself not to be the case, despite the fact that there is no obvious structure in the data. The non-zero value of $T_{Y_{\text{schuffled}}\to X}(m,l)$ must then be a byproduct of the finite data set. Definition (36) then ensures that spurious effects caused by finite m and l are removed.

4. Rényian transfer entropies

There are various ways in which one can sensibly define a transfer entropy with Rényi's information measure $S_q^{(R)}$. The most natural definition is the one based on a q-analog of Eqs. (32)–(33), i.e.,

$$\begin{aligned} T^{(R)}_{q;Y\to X}(m,l) &= S_q^{(R)}(X_{t_{m+1}}|X_{t_1}\cap\cdots\cap X_{t_m}) - S_q^{(R)}(X_{t_{m+1}}|X_{t_1}\cap\cdots\cap X_{t_m}\cap Y_{t_1}\cap\cdots\cap Y_{t_l}) \\ &= I_q^{(R)}(X_{t_{m+1}}; Y_{t_1}\cap\cdots\cap Y_{t_l}|X_{t_1}\cap\cdots\cap X_{t_m}). \end{aligned} \tag{37}$$

With the help of (25) and (30), this can be written in an explicit form as

$$\begin{aligned} T^{(R)}_{q;Y\to X}(m,l) &= \frac{1}{1-q}\log_2 \frac{\sum \varrho_q(x_{t_1},\ldots,x_{t_m}) p^q(x_{t_{m+1}}|x_{t_1},\ldots,x_{t_m})}{\sum \varrho_q(x_{t_1},\ldots,x_{t_m},y_{t_{m-l+1}},\ldots,y_{t_m}) p^q(x_{t_{m+1}}|x_{t_1},\ldots,x_{t_m},y_{t_{m-l+1}},\ldots,y_{t_m})} \\ &= \frac{1}{1-q}\log_2 \frac{\sum \varrho_q(x_{t_1},\ldots,x_{t_m}) p^q(y_{t_{m-l+1}},\ldots,y_{t_m}|x_{t_1},\ldots,x_{t_m})}{\sum \varrho_q(x_{t_1},\ldots,x_{t_{m+1}}) p^q(y_{t_{m-l+1}},\ldots,y_{t_m}|x_{t_1},\ldots,x_{t_{m+1}})}. \end{aligned} \tag{38}$$

Here, ϱ_q is the escort distribution (15). One can again easily check that in the limit $q \to 1$ we regain the Shannonian transfer entropy (34).

The representation (38) deserves a few comments. First, when the history of Y up to time t_m has no influence on the next-time-tick value of X (i.e., on $X_{t_{m+1}}$), then from the first line in (38) it follows that $T^{(R)}_{q;Y\to X}(m, l) = 0$, which indicates that no information flows from Y to X, as should be expected. In addition, $T^{(R)}_{q;Y\to X}$ as defined by (37) and (38) takes into account only the effect of time series Y (up to time t_m), while the compound effect of the time series X (up to time t_m) is subtracted (though indirectly present via correlations that exist between time series X and Y). In the spirit of Section 2.2 one may interpret the transfer entropy $T^{(R)}_{q;Y\to X}$ as a *rating factor* which quantifies a gain/loss in the risk concerning the behavior of X at the future time t_{m+1} after we take into account the historical values of a time series Y until t_m.

Unlike in Shannon's case, $T^{(R)}_{q;Y\to X} = 0$ does not imply independence of the X and Y processes. This is because $T^{(R)}_{q;Y\to X}(m, l)$ can also be *negative* on account of nonlinear pricing. Negativity of $T^{(R)}_{q;Y\to X}$ then simply means that the knowledge of historical values of both X and Y broadens the tail part of the anticipated PDF for the price value $X_{t_{m+1}}$ more than historical values of X only would do. In other words, extra knowledge of historical values of Y reveals a greater risk in the next time step of X than one would anticipate by knowing merely the historical data of X alone.

Note that, with our definition (37), $T_{q;Y\to X}$ is again explicitly directional since it measures the degree of dependence of X on Y and not the other way around, though in this case we should indicate by an arrow whether the original risk rate about $X_{t_{m+1}}$ was increased or reduced by observing the historical values of Y.

At this stage, one may introduce the effective Rényi transfer entropy (ERTE) by following the same logic as in the Shannonian case. In particular, one can again use the surrogate data technique to define the ERTE as

$$T^{(R,\mathrm{eff})}_{q;Y\to X}(m, l) \equiv T^{(R)}_{q;Y\to X}(m, l) - T^{(R)}_{q;Y_{\mathrm{schuffled}}\to X}(m, l). \tag{39}$$

Similarly to the RTE, $T^{(R)}_{q;Y\to X}(m, l)$ also accentuates for $q \in (0, 1)$ the flow of information that exists between the tail parts of distributions; i.e., it describes how marginal events in the time series Y influence marginal events in the time series X. Since most of historical data belong to the central parts of distributions (typically with well-behaved Gaussian increments), one can reasonably expect that for $q \in (0, 1)$ the transfer entropy $T^{(R,\mathrm{eff})}_{q;Y\to X}(m, l) \cong T^{(R)}_{q;Y\to X}(m, l)$, and the surrogate data technique is not needed. This fact is indeed confirmed in our data analysis presented in the following section.

5. Presentation of the analyzed data

In the subsequent analysis, we use two types of data set to illustrate the utility of Rényi's transfer entropy. The first data set consists of 11 stock exchange indices, sampled at a daily (end of trading day) rate. The data set was obtained from Yahoo financial portal (historical data) with help of the R-code program [46] for the period of time between 2 January 1998 and 31 December 2009. These data will be used to demonstrate quantitatively the statistical coherence of all the mentioned indices in the form of *heat maps* and *net flows*.

Because we also wish to illustrate our approach quantitatively, we use as a second data set time series of 183,308 simultaneously recorded data points from two market indices, namely from the DAX index and the S&P500 index, gathered on a minute-tick basis in the period from 2 April 2008 to 11 September 2009. In our analysis, we use complete records, i.e., minute data where only valid values for both the DAX index and the S&P500 index are admitted: periods without trading activity (weekends, nighttime, holidays) in one or both stock exchanges were excluded. This procedure has the obvious disadvantage that records substantially separated in real time may become close neighbors in the newly defined time series. Fortunately, relatively the small number of such "critical" points compared to the regular ones prevents a statistically significant error. In addition, due to computer data massification one may reasonably expect that the trading activity responds almost immediately to external stimuli. For this reason we have synchronized the data series to an identical reference time, a master clock, which we take to be Central European Time (CET).

5.1. Numerical calculation of transfer entropies

In order to find the PDF involved in definitions (34) (respectively, (36)) and (38) (respectively, (39)) we use the relative-frequency estimate. For this purpose we divide the amplitude (i.e., stock index) axis into N discrete amplitude bins and assign to every bin a sample data value. The number of data points per bin divided by the length of the time series then constitutes the relative frequency which represents the underlying empirical distribution. In order to implement the R-code in ETE and ERTE calculations we partition the data into disjoint equidistant time intervals (blocks), which serve as a coarse-graining grid. The number of data points we employ in our calculations is constant in each block. In each block only the arithmetic mean price is considered in block-dependent computations.

It is clear that the actual calculations depend on the number of bins chosen (this is also known as the alphabet length). In Ref. [7], it was argued that, in large data sets such as our time series, the use of alphabets with more than a few symbols is not compatible with the amount of data at one's disposal. In order to make a connection with existing results (see Refs. [7,8]), we conduct calculations at fixed alphabet length $N = 3$ (see Tables 1–3).

Table 1
Numerical data for the ETE that are used to generate Figs. 5 and 6.

	GSPC	GDAXI	ATX	SSMI	AORD	BSESN	HSI	N225	DJA	NY	IXIC
GSPC	0,00000	0,64294	0,20563	0,58809	0,13525	0,13135	0,24831	0,12437	0,18488	0,31810	0,63655
GDAXI	0,77440	0,00000	0,12875	1,06841	0,23264	0,07227	0,39049	0,06604	0,62008	0,71447	0,95265
ATX	0,30296	0,17621	0,00000	0,20245	0,25843	0,31853	0,11564	0,08610	0,20097	0,29119	0,22579
SSMI	0,53800	0,69405	0,13986	0,00000	0,25384	0,14018	0,40735	0,05545	0,54607	0,49822	0,71238
AORD	1,36056	1,51079	0,65343	1,86269	0,00000	0,84022	1,46986	0,25226	1,70907	1,55052	1,89641
BSESN	0,08470	0,08102	0,10800	0,15650	0,12057	0,00000	0,18340	0,03929	0,13135	0,07823	0,29994
HSI	0,32333	0,52722	0,21966	0,65140	0,28438	0,37382	0,00000	0,15690	0,39511	0,39078	0,79652
N225	0,98882	0,99773	0,85630	1,51051	0,84226	1,32050	0,83978	0,00000	1,32331	1,24970	1,08653
DJA	0,42696	0,77854	0,15381	0,95252	0,18801	0,14302	0,38341	0,08907	0,00000	1,75167	1,10667
NY	1,54040	0,58823	0,15845	0,57273	0,21266	0,11554	0,22785	0,10262	0,22374	0,00000	0,80568
IXIC	0,13514	0,23602	0,04650	0,24937	0,08140	0,06606	0,17859	0,04118	0,17067	0,18350	0,00000

Table 2
Numerical data for the ERTE that are used to generate Figs. 7 and 8; $q = 1.5$.

	GSPC	GDAXI	ATX	SSMI	AORD	BSESN	HSI	N225	DJA	NY	IXIC
GSPC	0,00000	0,75542	0,23044	0,68704	0,18199	0,14466	0,29378	0,13147	0,20075	0,29827	0,70910
GDAXI	0,87416	0,00000	0,16289	1,26024	0,29135	0,06898	0,46002	0,06911	0,74582	0,80991	1,06154
ATX	0,29124	0,17517	0,00000	0,20228	0,25199	0,31975	0,10526	0,08322	0,20047	0,28042	0,21941
SSMI	0,66533	0,83485	0,14882	0,00000	0,30681	0,16143	0,48324	0,06393	0,68161	0,61874	0,88317
AORD	1,43285	1,63308	0,68199	2,00807	0,00000	0,88969	1,56056	0,23869	1,75402	1,65517	1,96782
BSESN	0,08182	0,08539	0,09824	0,17994	0,13056	0,00000	0,22778	0,03927	0,14104	0,07924	0,33590
HSI	0,39850	0,60735	0,20366	0,78253	0,33898	0,45238	0,00000	0,15589	0,48579	0,47254	1,01988
N225	0,99591	1,00682	0,87902	1,50917	0,85318	1,33146	0,83274	0,00000	1,33389	1,25635	1,10432
DJA	0,45633	0,85978	0,16292	1,10406	0,24172	0,15696	0,44653	0,09256	0,00000	1,11094	1,15357
NY	-1,51108	0,69553	0,17910	0,66445	0,25290	0,13395	0,26325	0,10823	0,26325	0,00000	0,83430
IXIC	0,15980	0,31494	0,05382	0,34622	0,12063	0,09491	0,27096	0,04925	0,22103	0,21860	0,00000

Table 3
Numerical data for the ERTE that are used to generate Figs. 9 and 10; $q = 0.8$.

	GSPC	GDAXI	ATX	SSMI	AORD	BSESN	HSI	N225	DJA	NY	IXIC
GSPC	0,00000	0,59988	0,19689	0,55001	0,11760	0,12640	0,23150	0,12188	0,18844	0,35837	0,59674
GDAXI	0,73595	0,00000	0,11610	0,99500	0,21072	0,07367	0,36417	0,06500	0,57206	0,67796	0,91007
ATX	0,30781	0,17697	0,00000	0,20278	0,26107	0,31832	0,11974	0,08729	0,20142	0,29575	0,22852
SSMI	0,48954	0,64025	0,13663	0,00000	0,23401	0,13220	0,37856	0,05228	0,49472	0,45253	0,64689
AORD	1,33377	1,46457	0,64249	1,80720	0,00000	0,82122	1,43513	0,25782	1,69300	1,51069	1,87012
BSESN	0,08593	0,07945	0,11216	0,14773	0,11678	0,00000	0,16665	0,03931	0,13135	0,07805	0,28653
HSI	0,29513	0,49715	0,22674	0,60207	0,26371	0,34477	0,00000	0,15772	0,36080	0,35985	0,71247
N225	0,98648	0,99462	0,84811	1,51145	0,83787	1,31649	0,84356	0,00000	1,31936	1,24723	1,07973
DJA	0,44051	0,74738	0,15077	0,89469	0,16760	0,13769	0,35985	0,08787	0,00000	-2,57465	1,08989
NY	-1,98440	0,54747	0,15095	0,53761	0,19733	0,10858	0,21460	0,10079	0,21396	0,00000	0,80749
IXIC	0,12602	0,20599	0,04414	0,21223	0,06682	0,05549	0,14385	0,03814	0,14862	0,17055	0,00000

For a given partition, i.e., fixed l, $T_{Y\to X}(m, l)$ is a function of the block length m. The parameter m is to be chosen as large as possible in order to find a stable (i.e., large m independent) value for $T_{Y\to X}(m, l)$; however, due to the finite size of the real time series X, it is required to find a reasonable compromise between unwanted finite sample effects and a high value for m. This is achieved by substituting $T_{Y\to X}(m, l)$ with the effective transfer entropy. Surrogate data that are needed in definitions of the ETE (36) and the ERTE (39) are obtained by means of standard R routines [46]. The effective Rényi and Shannon transfer entropies themselves are explicitly calculated and visualized with the help of the open-source statistical framework R and its related R packages for graphical presentations. The calculations themselves are also coded in the R language.

5.2. Analyzing the daily data—heat maps versus net information flows

The effective transfer entropies $T^{\text{eff}}_{Y\to X}$ and $T^{(R,\text{eff})}_{q;Y\to X}$ are calculated between 11 major stock indices (see the list in Appendix A). The results are collected in three tables in Appendix B and applied in the constructions of *heat maps* and *net information flows* in Figs. 4–10. In particular, Shannon's information flow is employed in Figs. 5 and 6, while Rényi's transfer entropy is used in construction of Figs. 7–10. The histogram-based heat map in Fig. 4 represents the overall run of the 11 aforementioned indices after the filtering procedure. We have used the RColorBrewer package [47] from the R statistical environment which employs a color-spectrum visualization for asset prices. In this case the color runs from the green, for higher prices, to dark purple, for low price values.

The heat map in Fig. 5 shows that among the 11 selected markets a substantial amount of information flows between the Asia–Pacific region (APR) and the US. One can also clearly recognize the strong information exchanges between the APR and

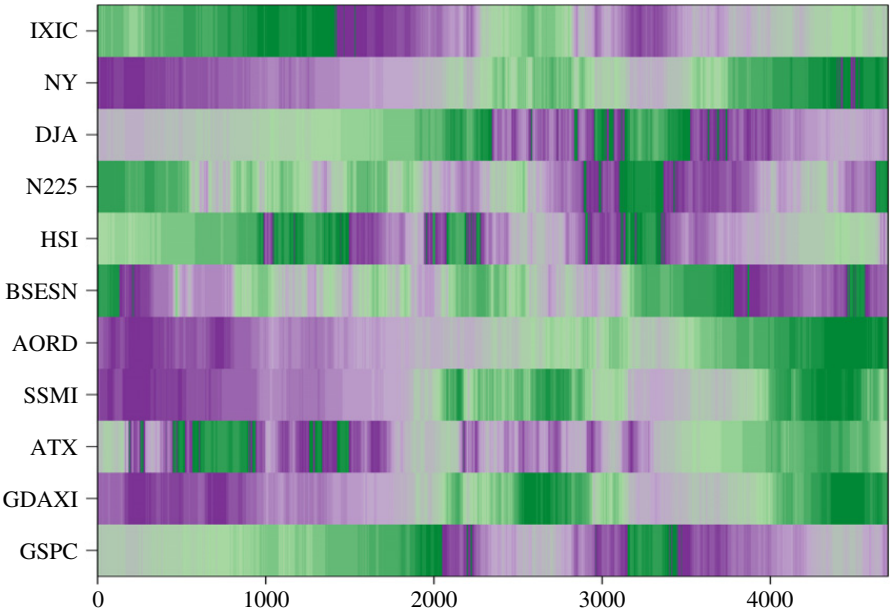

Fig. 4. Histogram-based (i.e., non-entropic) heat map between the 11 stock indices listed in Appendix A.

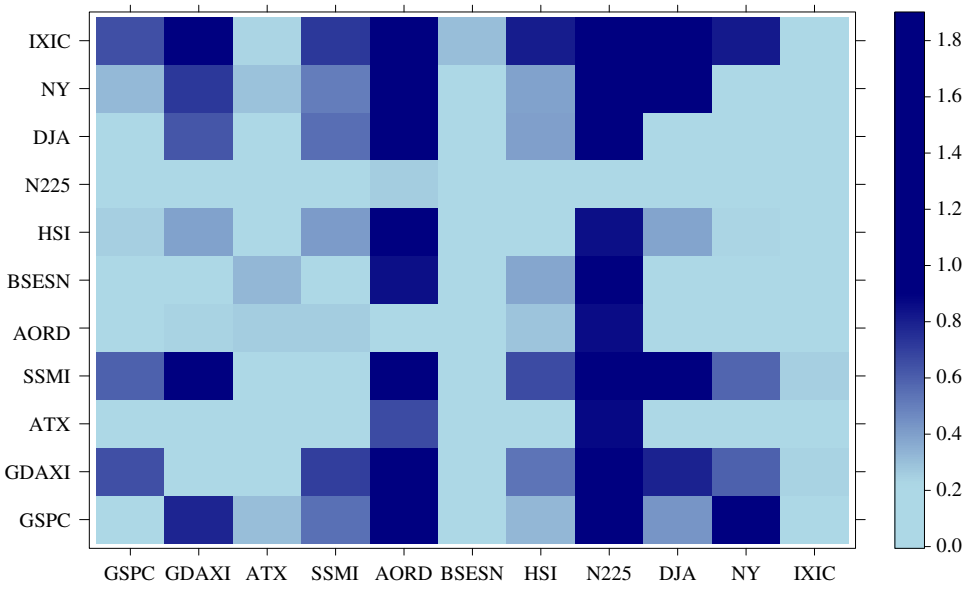

Fig. 5. Heat map of Shannon's effective entropy between the 11 stock indices listed in Appendix A. Alphabet size $N = 3$.

European markets and the subdominant information flow between the US and Europe. There is comparably less information flowing among European markets themselves. This can be credited to the fact that the internal European market is typically liquid and well equilibrated; similarly, a system in thermal equilibrium (far from critical points) has very little information flow among various parts. An analogous pattern (save for the NY index) can also be observed among the US markets. In contrast, the markets within the APR mutually exchange a relatively large volume of information. This might be attributed to a lower liquidity and consequently less balanced internal APR market.

The heat maps in Figs. 7 and 9 bring further understanding. Notably, we can see that the information flow within APR markets is significantly more imbalanced between wings of the asset distributions (larger color fluctuations) than between the corresponding central parts. This suggests low liquidity risks. A similar though subordinate imbalance in the information transfer can also be observed between the US and APR markets.

Understandably more revealing are the net information flows presented in Figs. 6, 8 and 10. The net flow $F_{Y\leftrightarrow X}$ is defined as $F_{Y\leftrightarrow X} \equiv T_{Y\to X} - T_{X\to Y}$. This allows one to visualize more transparently the disparity between the $Y \to X$ and $X \to Y$ flows. For instance, in Fig. 6 we see that substantially more information flows from the APR to the US and Europe than vice versa. Figs. 8 and 10 then demonstrate more specifically that the APR $\to$ Europe flow is evenly distributed between the central and tail distribution parts. From the net flow in Figs. 6, 8 and 10 we can also observe an important yet comparably weaker

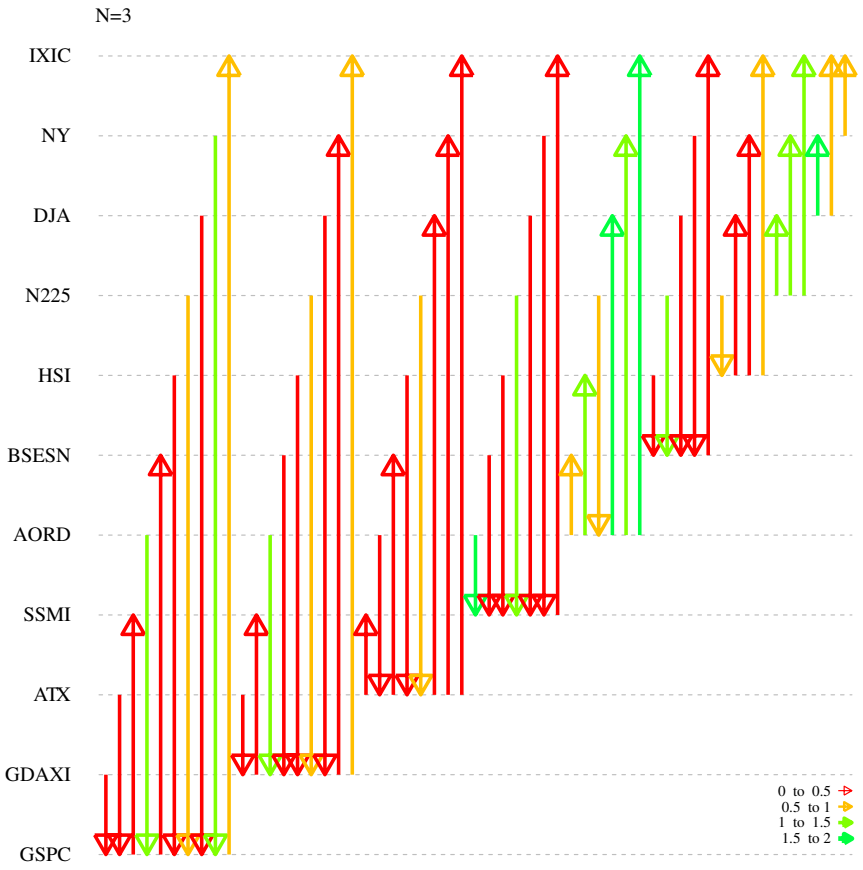

Fig. 6. Net flow $F_{Y\leftrightarrow X}$ of effective Shannon transfer entropies between the 11 stock indices listed in Appendix A. Alphabet size $N = 3$.

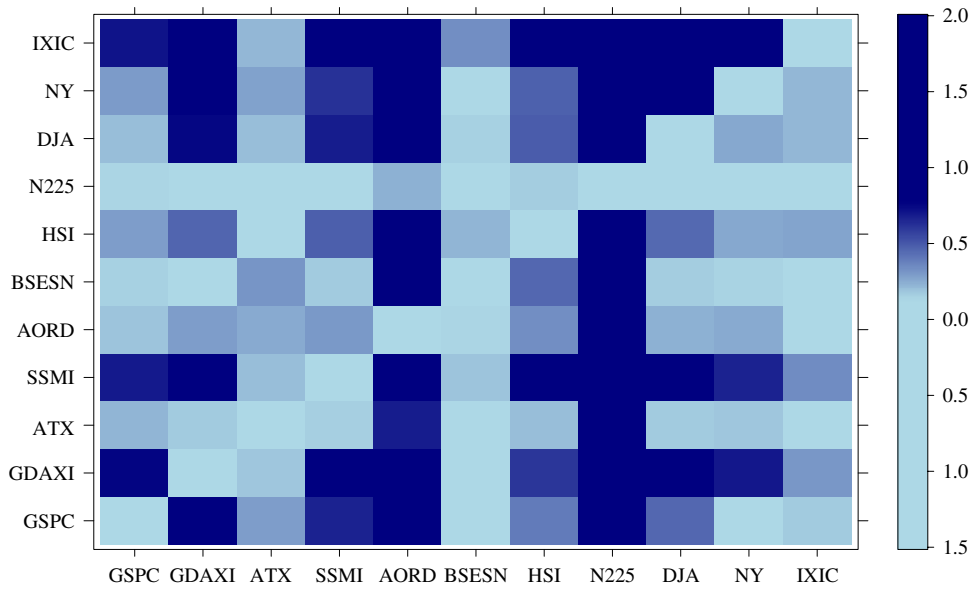

Fig. 7. Heat map of Rényi's effective entropy between the 11 stock indices listed in Appendix A; $q = 1.5$. Alphabet size $N = 3$.

surplus of information flow from Europe towards the US. This interesting fact will be further addressed in the following subsection.

Note also that $T^{(R)}_{1.5;\mathrm{SP\&500}\to\mathrm{NY}}$, $T^{(R)}_{0.8;\mathrm{SP\&500}\to\mathrm{NY}}$ and $T^{(R)}_{0.8;\mathrm{NY}\to\mathrm{DJ}}$ have negative values. These exceptional behaviors can be partly attributed to the fact that both the SP&500 and DJ indices are built from indices that are also present in the NY index and hence one might expect unusually strong coherence between these indices. From Section 4 we know that negative values of the ERTE imply a higher risk involved in a next-time-step asset-price behavior than could be predicted (or expected)

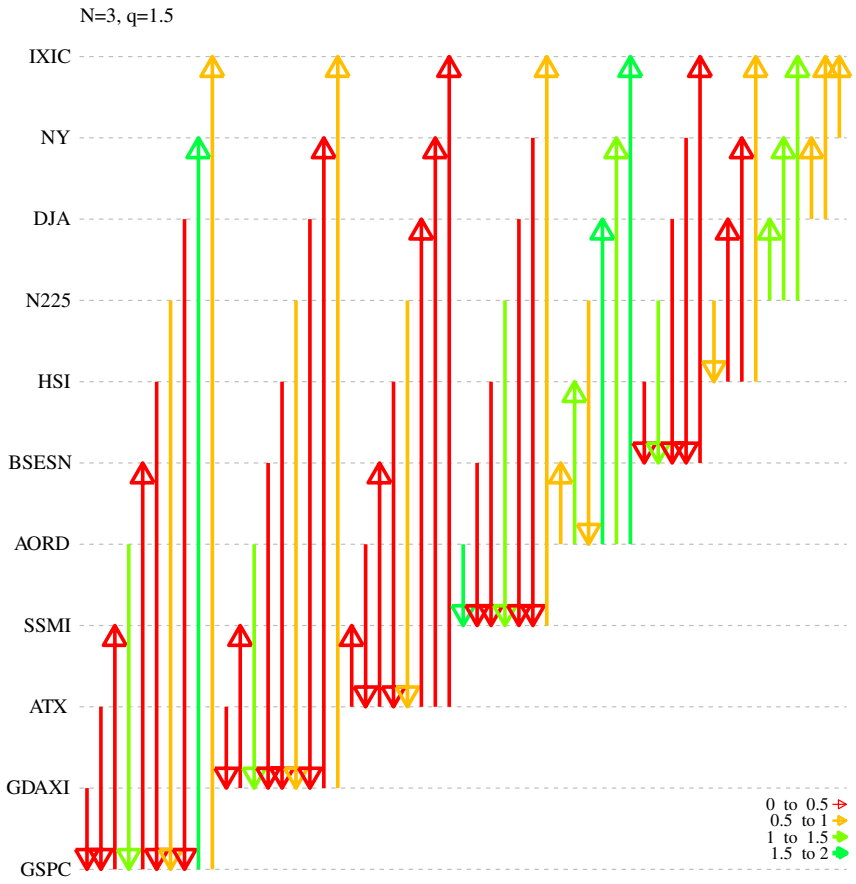

Fig. 8. Net flow $F_{Y\leftrightarrow X}$ of effective Rényi transfer entropies between the 11 stock indices listed in Appendix A; $q = 1.5$. Alphabet size $N = 3$.

without knowing the historical values of the source time series. The observed negativity of $T^{(R)}_{0.8;X\to Y}$ thus means that when some of the ignorance is elevated by observing the time series X a higher risk reveals itself in the nearest-future behavior of the asset price Y. Analogously, negativity of $T^{(R)}_{1.5;X\to Y}$ corresponds to a risk enhancement of the non-risky (i.e. close-to-peak) part of the underlying PDF.

5.3. Minute-price information flows

Here we analyze the minute-tick historic records of the DAX and S&P500 indices collected over the period of 18 months from 2 April 2008 to 11 September 2009. The coarse-grained overall run of both indices after the filtering procedure is depicted in the histogram-based heat map in Fig. 4.

Without any a prior knowledge about the Markovian (or non-Markovian) nature of the data series, we consider the order of the Markov process for both the DAX and S&P500 stocks to be identical, i.e., the price memory of both indices is considered to be the same. The latter may be viewed as a "maximally unbiased" assumption. At this stage we eliminate the surrogate data and consider the RTE alone. The corresponding RTEs for $q = 1.5$ and $q = 0.8$ as functions of block lengths are shown in Figs. 11 and 12, respectively. There we can clearly recognize that for $m \sim 200$–300 minutes there are no new correlations between the DAX and S&P500 indices. So, the underlying Markov process has order (or memory) roughly 200–300 min.

The aforementioned result is quite surprising in view of the fact that autocorrelation functions of stock market returns typically decay exponentially with a characteristic time of the order of minutes (e.g., ∼4 min for the S&P500 [48,49]), so the returns are basically uncorrelated random variables. Our result, however, indicates that two markets can be intertwined for much longer. This situation is actually not so surprising when we realize that empirical analysis of financial data asserts (see, e.g., Ref. [50]) that autocorrelation functions of higher-order correlations for asset returns have longer decorrelation time, which might span up to years (e.g., a few months in the case of volatility for the S&P500 [49]). It is indeed a key advantage of our approach that the *nonlinear* nature of the RTE naturally allows one to identify the existing long-time cross-correlations between financial markets.

In Fig. 13, we depict the empirical dependence of the ERTE on the parameter q. Despite the fact that the RE itself is a monotonically decreasing function of q (see, e.g., Ref. [26]) this is generally not the case for the ERTE (nor for the conditional RE). Indeed, the ERTE represents a difference of two REs with identical q (see Eq. (37)), and as such it may be neither monotonic nor decreasing. The functional dependence of the ERTE on q nevertheless serves as an important indicator of how quickly the REs involved change with q.

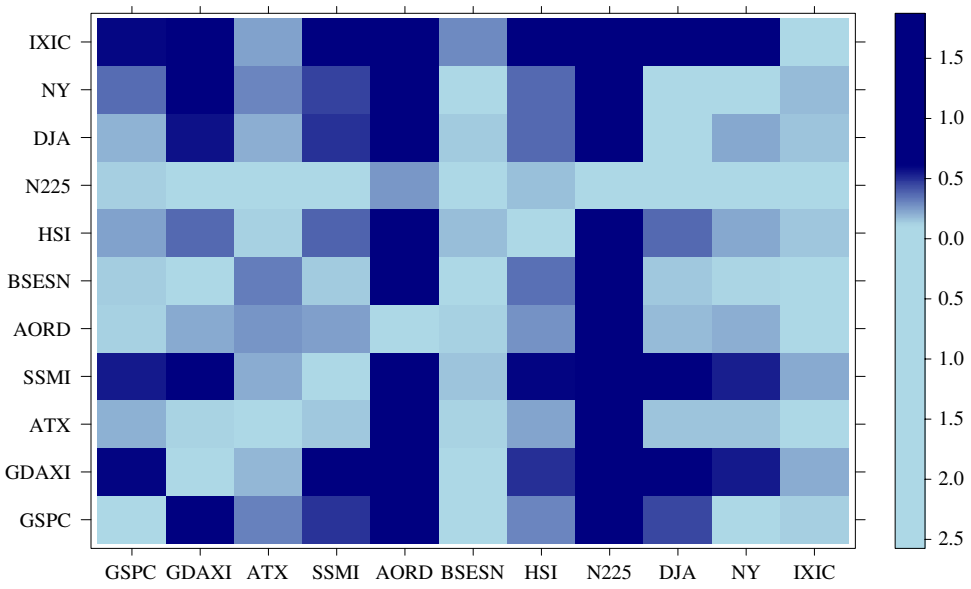

Fig. 9. Heat map of Rényi's effective entropy between the 11 stock indices listed in Appendix A; $q = 0.8$. Alphabet size $N = 3$.

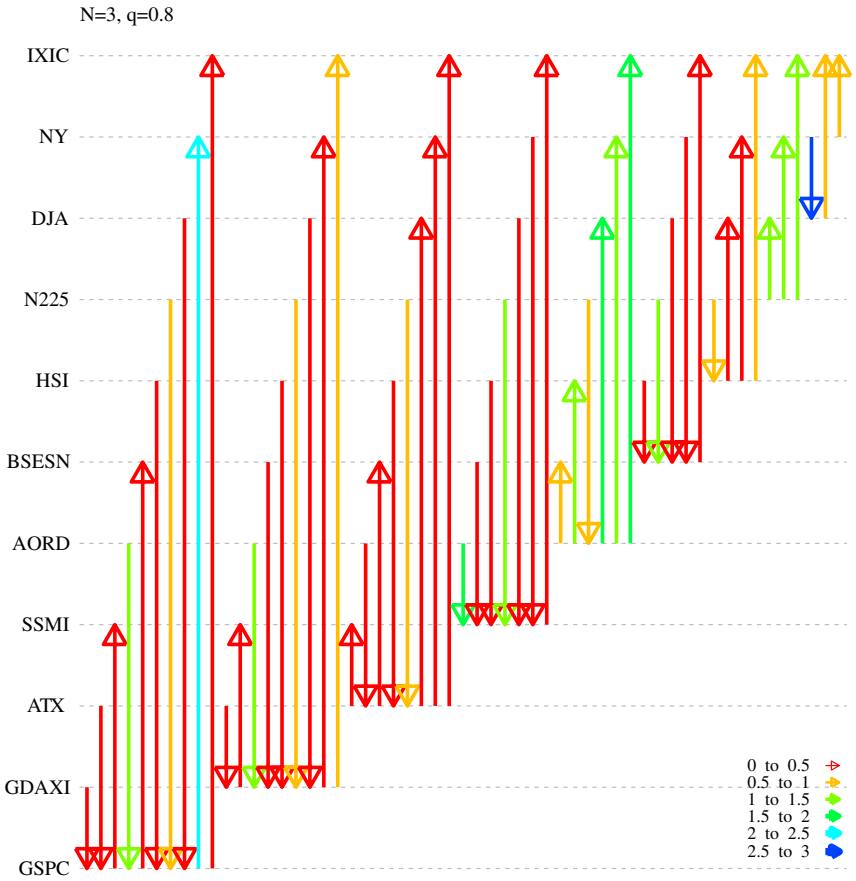

Fig. 10. Net flow $F_{Y\leftrightarrow X}$ of the effective Rényi transfer entropies between the 11 stock indices listed in Appendix A; $q = 0.8$. Alphabet size $N = 3$.

The results reproduced in Fig. 13 quantitatively confirm the expected asymmetry in the information flow between the US and European markets. However, since the US contributes more than half of the world's trading volume, it could be anticipated that there is a stronger information flow from big US markets towards both European and APR markets. Yet, despite the strong US trading record, our ERTE approach indicates that the situation is not so straightforward when the entropy-based information flow is considered as a measure of market cross-correlation. Indeed, from Figs. 6, 8 and 10 we could observe that there is a noticeably stronger information flow from the European and APR markets to the US markets

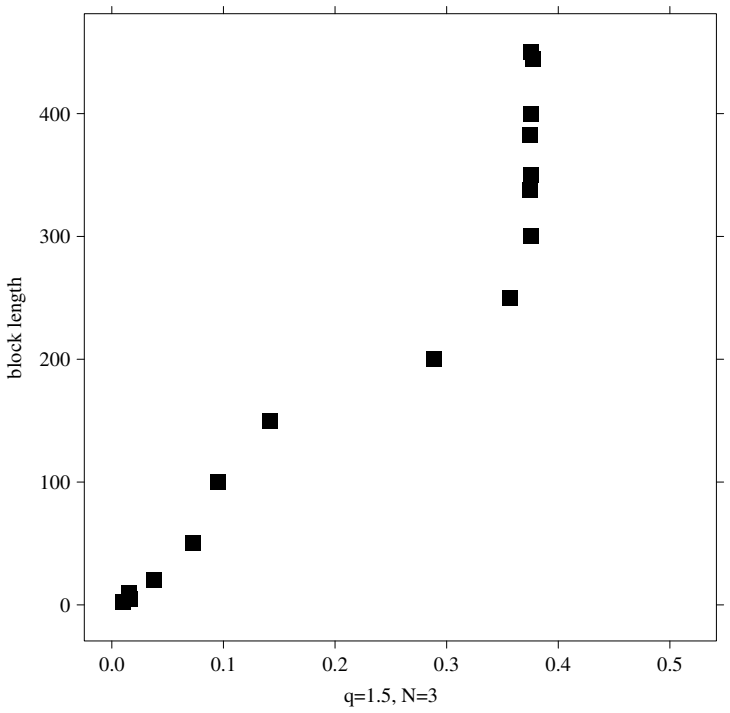

Fig. 11. Effective Rényi transfer entropy $T^{(R)}_{1.5;\mathrm{SP\&500}\to\mathrm{DAX}}(m, m)$ for a 3-letter alphabet as a function of the block length m. DAX and SP&500 min prices are employed. The correlation time is between 200–300 minutes.

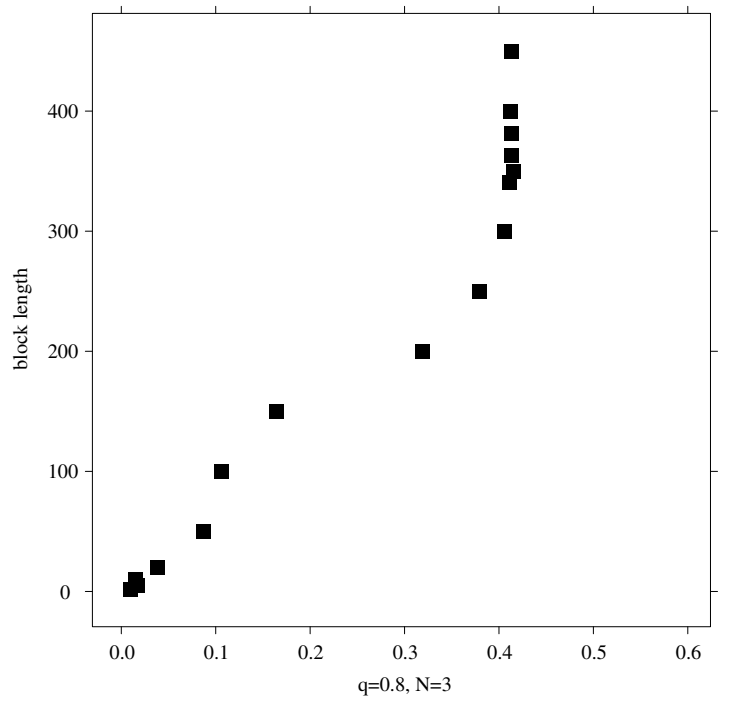

Fig. 12. Effective Rényi transfer entropy $T^{(R)}_{0.8;\mathrm{SP\&500}\to\mathrm{DAX}}(m, m)$ for a 3-letter alphabet as a function of the block length m. DAX and SP&500 min prices are employed. The correlation time is between 200–300 minutes.

than vice versa. Fig. 13 extends the validity of this observation to short time scales of the order of minutes. In particular, from Fig. 13 we clearly see that flow from the DAX to the S&P500 is stronger than the reverse flow. It is also worth noting that this Europe–US flow is positive for all values of q, i.e., for all distribution sectors, with a small bias towards tail parts of the underlying distribution.

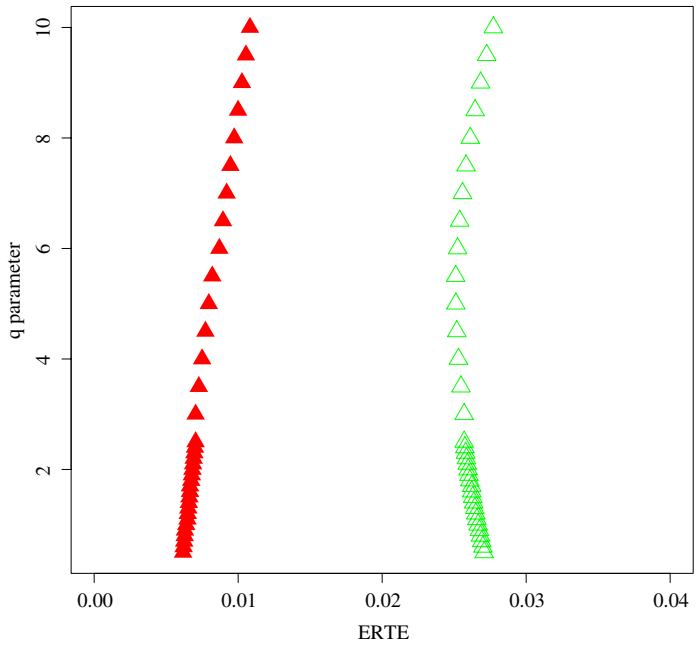

Fig. 13. The ERTE as a function of q. The alphabet size is set to $N = 3$. DAX and SP&500 min prices are employed. The red curve corresponds to $T^{(R,\mathrm{eff})}_{q;\mathrm{SP\&500}\to\mathrm{DAX}}(m, m)$ while the green curve denotes $T^{(R,\mathrm{eff})}_{q;\mathrm{DAX}\to\mathrm{SP\&500}}(m, m)$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

6. Concluding remarks

Transfer entropies have been repeatedly utilized in the quantification of statistical coherence between various time series with prominent applications in financial markets. In contrast to previous works in which transfer entropies have been exclusively considered only in the context of Shannon's information theory, we have advanced here the notion of Rényi's (i.e. non-Shannonian) transfer entropy. The latter is defined in a close analogy with Shannon's case, i.e., as the information flow (in bits) from Y to X ignoring static correlations due to the common historical factors such as external agents or forces. However, unlike Shannon's transfer entropy, where the information flow between two (generally cross-correlated) stochastic processes takes into account the whole underlying empirical price distribution, the RTE describes the information flow only between certain pre-decided parts of two price distributions involved. The distribution sectors in question can be chosen when Rényi's parameter q is set in accordance with Campbell's pricing theorem. Throughout this paper we have demonstrated that the RTE thus defined has many specific properties that are desirable for the quantification of an information flow between two interrelated stochastic systems. In particular, we have shown that the RTE can serve as an efficient rating factor which quantifies a gain or loss in the risk that is inherent in the passage from X_{t_m} to $X_{t_{m+1}}$ when new information, namely historical values of a time series Y until time t_m, is taken into account. This gain/loss is parameterized by a single parameter, the Rényi q parameter, which serves as a "zooming index" that zooms (or emphasizes) different sectors of the underlying empirical PDF. In this way one can scan various sectors of the price distribution and analyze the associated information flows. In particular, the fact that one may separately scrutinize information fluxes between tails or central-peak parts of asset price distributions simply by setting $q < 1$ or $q > 1$, respectively, can be employed, for example, by financial institutions to quickly analyze the global (across-the-border) information flows and use them to redistribute their risk. For instance, if an American investor observes that a certain market, say the S&P500, is going down and he/she knows that the corresponding NASDAQ ERTE for $q < 1$ is low, then he/she does not need to relocate the portfolio containing related assets rapidly, because the influence is in this case slow. Slow portfolio relocation is generally preferable, because fast relocations are always burdened with excessive transaction costs. Let us stress that this type of conduct could not be deduced from Shannon's transfer entropy alone. In fact, the ETE suggests a fast (and thus expensive) portfolio relocation as a best strategy (see Figs. 6, 8 and 10).

Let us stress that applications of transfer entropies presented quantitatively support the observation that more information flows from the Asia–Pacific region towards the US and Europe than vice versa, and this holds for transfers between both peak parts and wing parts of asset PDFs; i.e., the US and European markets are more prone to price shakes in the Asia–Pacific sector than the other way around. Besides, information-wise the US market is more influenced by the European one than in reverse. This interesting observation can be further substantiated by our DAX versus S&P500 analysis, in which we have seen that the influx of information from Europe is to a large extent due to a tail-part transfer. The peak-part transfer is less pronounced. So, although the US contributes more than half of the world's trading volume, our results indicate that this is not the case with information flow. In fact, the US markets seem to be prone to reflect a marginal (i.e.,

risky) behavior in both European and APS markets. Such a fragility does not seem to be reciprocated. This point definitely deserves further closer analysis.

Finally, one might be interested in how the RTE presented here compares with other correlation tests. The usual correlation tests take into account either the lower-order correlations (e.g., time-lagged cross-correlation test and Arnhold et al. interdependence test) or they try to address the causation issue between bivariate time series (e.g., Granger causality test or Hacker and Hatemi-J causality test). Since the RTE allows one to compare only certain parts of the underlying distributions it also works implicitly with high-order correlations, and for the same reason it cannot affirmatively answer the causation issue. In many respects such correlation tests bring complementary information with respect to the RTE approach. More detailed discussion concerning multivariate time series and related correlation tests will be presented elsewhere.

Acknowledgments

This work was partially supported by the Ministry of Education of the Czech Republic (research plan MSM 6840770039), and by the Deutsche Forschungsgemeinschaft under grant Kl256/47.

Appendix A

In this appendix we provide a brief glossary of the indices used in the main text. The notation presented here conforms with the notation typically listed in various on-line financial portals (e.g., Yahoo financial portal).

Indices	Description	Country
GSPC (S&P500)	Standard and Poor 500 (500 stocks actively traded in the US).	USA
GDAXI (DAX)	Dax Indices (stock of 30 major German companies).	Germany
ATX	The Austrian Traded Index is the most important stock market index of the Wiener Börse. The ATX is a price index and currently consists of 20 stocks.	Austria
SSMI	The Swiss Market Index is a capitalization-weighted index of the 20 largest and most liquid stocks. It represents about 85% of the free-float market capitalization of the Swiss equity market.	Swiss
AORD	All Ordinaries represents the 500 largest companies in the Australian equities market. Index constituents are drawn from eligible companies listed on the Australian Stock Exchange.	Australia
BSESN	The BSE Sensex is a market capitalized index that tracks 30 stocks from the Bombay Stock Exchange. It is the second largest exchange of India in terms of volume and first in terms of shares listed.	India
HSI	The Hang Seng Index denoted in Hong Kong stock market. It is used to record and monitor daily changes of the largest companies of the Hong Kong stock market. It consist of 45 Companies.	Hong Kong
N225	Nikkei 225 is a stock market index for the Tokyo Stock Exchange. It is a price-weighted average (the unit is yen), and the components are reviewed once a year. Currently, the Nikkei is the most widely quoted average of Japanese equities, similar to the Dow Jones Industrial Average.	Japan
DJA (DJ)	The Dow Jones Industrial Average, also referred to as the Industrial Average, the Dow Jones, the Dow 30, or simply as the Dow; it is one of several US stock market indices. First published in 1887.	USA
NY	iShares NYSE 100 Index is an exchange trading fund, which is a security that tracks a basket of assets, but trades like a stock. NY tracks the SE US 100; this equity index measures the performance of the largest 100 companies listed on the New York Stock Exchange (NYSE).	USA
IXIC (NASDAQ)	The Nasdaq Composite is a stock market index of all of the common stocks and similar securities (e.g., ADRs, tracking stocks, limited partnership interests) listed on the NASDAQ stock market, it has over 3000 components.	USA

Appendix B

In this appendix we specify explicit values of effective transfer entropies that are employed in Section 5. These are calculated for alphabet with $N = 3$ (see Tables 1–3).

References

[1] J.-S. Yang, W. Kwak, T. Kaizoji, I.-M. Kim, Eur. Phys. J. B 61 (2008) 389;
K. Matal, M. Pal, H. Salunkay, H.E. Stanley, Europhys. Lett. 66 (2004) 909;
H.E. Stanley, L.A.N. Amaral, X. Gabaix, P. Gopikrishnan, V. Plerou, Physica A 299 (2001) 1.
[2] J.-S. Yang, S. Chae, W.-S. Jung, H.-T. Moon, Physica A 363 (2006) 377.
[3] K. Kim, S.-M. Yoon, Physica A 344 (2004) 272.
[4] J.B. Park, J.W. Lee, J.-S. Yang, H.-H. Jo, H.-T. Moon, Physica A 379 (2007) 179.
[5] J.W. Lee, J.B. Park, H.-H. Jo, J.-S. Yang, H.-T. Moon, Physics, 0607282, 2006.
[6] T. Schreiber, Phys. Rev. Lett. 85 (2000) 461.
[7] R. Marschinski, H. Kantz, Eur. Phys. J. B 30 (2002) 275.
[8] O. Kwon, J.-S. Yang, Europhys. Lett. 82 (2008) 68003.
[9] M. Paluš, M. Vejmelka, Phys. Rev. E 75 (2007) 056211.
[10] M. Lungarella, A. Pitti, Y. Kuniyoshi, Phys. Rev. E 76 (2007) 056117.
[11] P. Jizba, T. Arimitsu, Ann. Phys. (NY) 312 (2004) 17.
[12] K. Huang, Statistical Mechanics, John Wiley & Sons, New York, 1963.
[13] The maximal entropy principle was introduced in statistical thermodynamics by J.W. Gibbs under the name "The fundamental hypothesis of equal a priori probabilities in the phase space".
[14] C.E. Shannon, Bell Syst. Tech. J. 27 (1948) 379. 623.
[15] C.E. Shannon, W. Weaver, The Mathematical Theory of Communication, University of Illinois Press, New York, 1949.
[16] A. Rényi, A Diary on Information Theory, John Wiley & Sons, New York, 1984.
[17] R. Ash, Information Theory, Wiley, New York, 1965.
[18] A.I. Khinchin, Mathematical Foundations of Information Theory, Dover Publications, Inc., New York, 1957.
[19] A. Feinstein, Foundations of Information Theory, McGraw Hill, New York, 1958.
[20] T.W. Chaudy, J.B. McLeod, Edinburgh Mathematical Notes, 43, 1960, p. 7;
for H. Tverberg's, P.M. Lee's or D.G. Kendall's axiomatics of Shannon's entropy, see, e.g., S. Guias, Information Theory with Applications, McGraw Hill, New York, 1977.
[21] L. Szilard, Z. Phys. 53 (1929) 840.
[22] L. Brillouin, J. Appl. Phys. 22 (1951) 334.
[23] E.T. Jaynes, Papers on Probability and Statistics and Statistical Physics, D. Reidel Publishing Company, Boston, 1983.
[24] I. Csiszár, P.C. Shields, Information Theory and Statistics: A Tutorial, Now Publishers Inc., Boston, 2004.
[25] A. Rényi, Probability Theory, North-Holland, Amsterdam, 1970.
[26] A. Rényi, Selected Papers of Alfred Rényi, vol. 2, Akademia Kiado, Budapest, 1976.
[27] L.L. Campbell, Inf. Control 8 (1965) 423.
[28] J. Aczél, Z. Daróczy, Measures of Information and their Characterizations, Academic Press, New York, 1975.
[29] J.-F. Bercher, Phys. Lett. A 373 (2009) 3235.
[30] This exponential weighting is also known as a Kolmogorov–Nagumo averaging. If the linear averaging is given by $\langle X\rangle = \sum_{x\in X}^{W} p(x)x$, the exponential weighting is defined as $\langle X\rangle_{\exp} = \phi^{-1}\left(\sum_{x\in X}^{W} p(x)\phi(x)\right)$, with $\phi(x) = 2^{\beta x}$. The $\beta = (1-q)/q$ factor is known as the Campbell exponent
[31] C. Beck, F. Schlögl, Thermodynamics of Chaotic Systems, Cambridge University Press, Cambridge, 1993.
[32] T.C. Halsey, M.H. Jensen, L.P. Kadanoff, I. Procaccia, B.I. Schraiman, Phys. Rev. A 33 (1986).
[33] E. Arikan, IEEE Trans. Inform. Theory 42 (1996) 99.
[34] F. Jelinek, IEEE Trans. Inform. Theory IT-14 (1968) 490.
[35] I. Csiszár, IEEE Trans. Inform. Theory 41 (1995) 26.
[36] Z. Daróczy, Acta Math. Acad. Sci. Hungaricae 15 (1964) 203.
[37] P. Jizba, T. Arimitsu, Physica A 340 (2004) 110.
[38] P. Jizba, T. Arimitsu, Physica A 365 (2006) 76.
[39] C. Tsallis, J. Stat. Phys. 52 (1988) 479. Bibliography URL: http://tsallis.cat.cbpf.br/biblio.htm.
[40] L. Golshani, E. Pasha, G. Yari, Inform. Sci. 179 (2009) 2426.
[41] C. Cahin, Entropy measures and unconditional security in cryptography, Ph.D. Thesis, Swiss Federal Institute of Technology, Zurich, 1997.
[42] E. Post, Trans. Amer. Math. Soc. 32 (1930) 723.
[43] One can map $S_q^{(R)}$ with $q \geq 1$ to $S_q^{(R)}$ with $0 < q \leq 1$ via duality $q \leftrightarrow 1/q$ that exists between ϱ_q and $\mathscr{P}$. In fact, we can observe that $S_{1/q}^{(R)}(\varrho_q) = S_q^{(R)}(\mathscr{P})$ and $S_q^{(R)}(\varrho_{1/q}) = S_{1/q}^{(R)}(\mathscr{P})$. So $S_q^{(R)}$ with $q \geq 1$ and $S_q^{(R)}$ with $0 < q \leq 1$ carry equal amount of information
[44] O. Kwona, J. Yanga, Physica A 387 (2008) 2851.
[45] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, J.D. Farmer, Physica D 58 (1992) 77.
[46] R Development Core Team, R: a language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, 2008. URL: http://www.R-project.org.
[47] E. Neuwirth, RColorBrewer: ColorBrewer.org Palettes, R package version 1.0-5, 2011. URl: http://CRAN.R-project.org/package=RColorBrewer.
[48] R.N. Mantegna, H.E. Stanley, An Introduction to Econophysics, Cambridge University Press, Cambridge, 2000.
[49] P. Jizba, H. Kleinert, P. Haener, Physica A 388 (2009) 3503.
[50] Y. Liu, P. Cizeau, M. Meyer, C.-K. Peng, H.E. Stanley, Physica A 245 (1997) 437; Physica A 245 (1997) 441;
Y. Liu, P. Gopikrishnan, P. Cizeau, M. Mayer, C.-K. Peng, H.E. Stanley, Phys. Rev. E 60 (1999) 1390.