Eur. Phys. J. B **30**, 275–281 (2002)
DOI: 10.1140/epjb/e2002-00379-2

THE EUROPEAN PHYSICAL JOURNAL B

Analysing the information flow between financial time series

An improved estimator for transfer entropy

R. Marschinski[1,2,a] and H. Kantz[1]

[1] Max-Planck-Institute for the Physics of Complex Systems, Nöthnitzer Str. 38, 01187 Dresden, Germany
[2] Dipartimento di Fisica, Università di Bologna, 40100 Bologna, Italy

Received 22 October 2001 / Received in final form 30 August 2002
Published online 29 November 2002 –

Abstract. Following the recently introduced concept of *transfer entropy*, we attempt to measure the information flow between two financial time series, the Dow Jones and DAX stock index. Being based on Shannon entropies, this model-free approach in principle allows us to detect statistical dependencies of all types, *i.e.* linear and nonlinear temporal correlations. However, when available data is limited and the expected effect is rather small, a straightforward implementation suffers badly from misestimation due to finite sample effects, making it basically impossible to assess the significance of the obtained values. We therefore introduce a modified estimator, called *effective transfer entropy*, which leads to improved results in such conditions. In the application, we then manage to confirm an information transfer on a time scale of one minute between the two financial time series. The different economic impact of the two indices is also recovered from the data. Numerical results are then interpreted on one hand as capability of one index to explain future observations of the other, and on the other hand within terms of coupling strengths in the framework of a bivariate autoregressive stochastic model. Evidence is given for a nonlinear character of the coupling between Dow Jones and DAX.

PACS. 02.50.-r Probability theory, stochastic processes, and statistics – 05.45.Tp Time series analysis – 89.90.+n Other topics in areas of applied and interdisciplinary physics

1 Introduction

Recently, Schreiber [1] introduced the information-theoretic inspired concept of transfer entropy, aimed at quantifying in a non-parametric and explicitly non-symmetric way the flow of information between two time series. Formally, this model-free approach allows to detect statistical dependencies in a very general way; in particular it is not limited to linear statistics, but potentially reveals all types of temporal correlations. However, due to relatively high amounts of data required – and especially when the expected effect is rather small – practical analyses are complicated by finite sample effects that make it difficult to assess the significance of obtained values. In the present work, we will therefore propose a slightly modified version of transfer entropy, to be called *effective transfer entropy* (ETE), that leads to an improved estimation under such conditions.

In order to illustrate the improved estimation by ETE in a practical case, we will test it by measuring the information flow between two historical financial time series, the US-American Dow Jones Industrial Average and the German DAX Xetra stock index. The numerical findings will be interpreted in a unique way within the framework of an artificial model based on a bivariate autoregressive stochastic process.

The motivation for such an analysis of financial data is by no means limited to testing purposes, since it has a definite value of its own. In fact, there has been a recent but growing interest of physicists in the dynamics of financial markets [2–7]. This might have been stimulated by the discovery of some strong analogies between speculative markets and some well known physical phenomena and concepts, as for instance spin systems [8], turbulence [3], universality [7], self-organised criticality [5,9], and complexity [6], almost all of which can be associated with the statistical mechanics branch of physics. With growing recognition of this new field of interest the term *econophysics* was being coined.

Today, one can divide the research activities within econophysics roughly in two areas: the "microscopic" approach investigates the financial market dynamics from the point of view of the single agents, with the long-term target of being able to reproduce the complex "macroscopic" behaviour of the financial markets starting from microscopic equations [5,8]. To thoroughly analyse and describe the statistical properties of that "macroscopic" behaviour is exactly what constitutes the second branch of econophysics [10–14]. This field of research profits from the immense amount of electronically recorded financial

[a] e-mail: `robert.marschinski@pik-potsdam.de`

data available; in that sense econophysics can also be considered a laboratory for methods of data analysis and, in particular and as in our case, for time series analysis.

So far, few authors within the econophysics community have studied the multivariate properties of financial time series: some works [4,6,7], applying *e.g.* random matrix theory, analyse the linear cross-correlations of stocks composing an exchange index; one paper [15] explicitly discusses cross-correlations between the Dow Jones and Dax index[1], but as do most such publications, it limits itself to the use of linear statistics. Information-theoretic tools, such as conditional entropies or mutual information, were applied only in few cases to financial data as for now [10,11,13].

2 Presentation of the data

The analysed dataset[2] consists of 63 867 simultaneously recorded data points of the Dow Jones (DJ) and DAX stock exchange indices, sampled at a one-minute rate, during the time between May 2000 and June 2001. Only complete records, *i.e.* minutes with a valid value for both DAX and DJ, were admitted: invalid values due to transmission errors or computer failures were carefully filtered out, and periods without trading activity (weekends, night-time, holidays) in one or both stock exchanges were excluded, reconnecting afterwards the remaining parts of the original time series. This procedure has the obvious drawback that records notedly separated in real time may become close neighbours in the newly defined time series, but the relatively small number of such "critical" points compared to the regular ones prevents a statistically significant impact[3]. The overall run of both indices after the preprocessing is displayed in Figure 1.

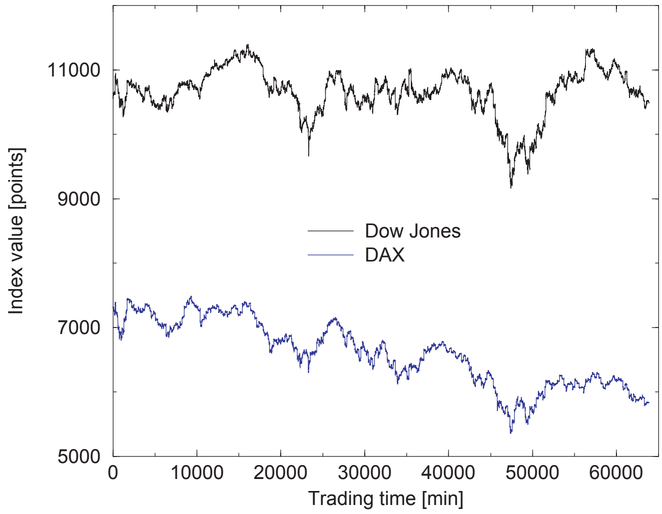

Fig. 1. Dow Jones (upper curve) and DAX: overall run of filtered time series.

As is evident from Figure 1, the raw DJ or DAX series cannot reasonably be assumed as stationary, a property yet essential for the validity of the forthcoming analysis. The standard solution to this problem is to define some new variable that can be considered sufficiently stationary, or at least asymptotically stationary [6]. The usual variables chosen by most authors to describe a financial time series $x(t), t = 1 \ldots N$ are the *price-change* or *increment*, $\delta x_\tau(t) := x(t+\tau) - x(t)$, *return*, $r_\tau(t) := \delta x_\tau(t)/x(t)$, or *log-return*, $s_\tau(t) := \ln[x(t+\tau)] - \ln[x(t)]$. The choice of the variable does not affect the outcome of the present work; in fact, in the high-frequency regime they are approximately identical, or proportional to each other [6]. Following the majority of authors, we will adopt the log-returns in our analysis. The usual quantity employed to characterise the fluctuation in financial data is the so called volatility, here defined as

$$vol_\Delta(t) := \frac{1}{\Delta} \sum_{i=1}^{\Delta} |s_\tau(t+i)|, \tag{1}$$

where the parameter Δ refers to the chosen length of the time-window and τ (in our case always $\tau = 1$ min) denotes the basic time scale. The average values of the DJ and DAX log-returns are $\langle \hat{s}(t)_{DJ} \rangle_t \simeq \langle \hat{s}(t)_{DAX} \rangle_t \simeq \pm 1 \times 10^{-6}$, while the absolute log-returns, also interpretable as an estimate of the one-minute volatility, have mean values of $\langle \hat{vol}_{1min}(t)_{DJ} \rangle_t \simeq \langle \hat{vol}_{1min}(t)_{DAX} \rangle_t \simeq 3 \times 10^{-4}$. However, as widely known, the strength of fluctuations in financial data is subject to long-term correlated oscillations[4]. Still, in concordance with other authors [6] we assume a sufficiently long financial time series to be asymptotically stationary, *i.e.* leading to relevant results for the large time statistical properties of the analysed data.

3 Definition of transfer entropy

Transfer entropy has been very recently introduced by Schreiber [1]. Its foundations are to be found in the works of Shannon [16] and Kolmogorov [17] on the theory of information [18]. Let us consider a discrete and stationary signal $I(t)$, with $p(i)$ being the probability[5] to observe symbol $i, i \in \{1, 2, \ldots, S\}$, and S denoting the number of symbols in the alphabet. According to Shannon, the average number of bits needed to optimally encode the signal I without taking into account possible correlations is given by

$$H_I := -\sum_{i=1}^{S} p(i) \log_2 p(i), \qquad 0 \le H_I \le \log_2 S, \tag{2}$$

called Shannon entropy. By writing $p(i_1, i_2, \ldots, i_m)$ for the probability of observing the subsequence $(i_1, i_2, \ldots, i_m)$,

[1] But only daily data was analysed.

[2] Provided by Deutsche Bank Research.

[3] In principle, the problem could be resolved by skipping these time indices when calculating $H_I(m)$.

[4] Known as correlated volatility.

[5] Time independent, since we assumed stationarity.

one can generalise the Shannon entropy and define the block-entropy of order m:

$$H_I(m) := -\sum_{i_1,\ldots,i_m=1}^{S} p(i_1, i_2, \ldots, i_m) \log_2 p(i_1, i_2, \ldots, i_m). \tag{3}$$

The differences of block-entropies of neighbouring order constitute the conditional entropies:

$$h_I(m) := H_I(m+1) - H_I(m), \qquad 0 \le h_I(m) \le H_I. \tag{4}$$

$h_I(m)$ expresses the average amount of information (in bits) still transmitted by the latest observation $I(m+1)$ when the last m observations of I are known and their information has been completely exploited; or, equivalently, the missing information for a correct forecast of $I(m+1)$ with the help of the m preceding historical observations. By using equation (3) and some elementary algebra, one can rewrite equation (4) as

$$h_I(m) = -\sum p(i_1, \ldots, i_{m+1}) \log_2 p(i_{m+1}|i_1, \ldots, i_m), \tag{5}$$

namely as Shannon entropy of the conditional probabilities, here denoting by

$$p(i_{m+1}|i_1, i_2, \ldots, i_m) = p(i_1, \ldots, i_m, i_{m+1})/p(i_1, \ldots, i_m) \tag{6}$$

the probability to observe symbol (i_{m+1}) immediately after the sequence $(i_1, i_2, \ldots, i_m)$. This also explains the name *conditional* entropy.

The limit $\lim_{m\to\infty} h_I(m) =: h_I$ takes the name *entropy of the source* and quantifies the average amount of information needed to predict a future observation when knowing the entire history of a series I. In case of a periodic signal one finds $h_I = 0$, $h_I = H_I$ for a purely stochastic and $0 < h_I < H_I$ for a chaotic or correlated signal [19]. In practice, probabilities are estimated through relative frequencies, $p(i_1, \ldots, i_m) = \frac{n(i_1,\ldots,i_m)}{N}$, where $n(i_1, \ldots, i_m)$ is the number of occurrences of the sequence $(i_1, \ldots, i_m)$ inside the data set, and N is the length of the time series. The limit $m \to \infty$ then is of course impossible to achieve, and therefore one must look for the asymptotic behaviour of $h_I(m)$, hoping to find a sufficiently large region with values independent of m (called plateau), before $h_I(m)$ gets substantially underestimated and eventually converges to zero due to finite sample effects.

Transfer entropy (TE) is closely related to conditional entropy, but it extends to two series, $I(t)$ and $J(t)$. The concept is the following:

Transfer Entropy =
+ information about future observation $I(t+1)$ gained from past joint observations of I and J
− information about future observation $I(t+1)$ gained from past observations of I only
= information flow from J to I.

This definition already reflects the key advantage of transfer entropy over other cross-correlation statistics: it is an asymmetric measure, that takes into account only statistical dependencies truly originating in the "source" series J, but not those deriving from a common history, like in the case of a common external drive for instance. Expressing the above relationship with the conditional entropies h_m and using equation (5) leads to

$$T_{J\to I}(m,l) := h_I(m) - h_{IJ}(m,l) \tag{7}$$

$$= \sum p(i_1, \ldots, i_{m+1}, j_1, \ldots, j_l) \times \log_2 \frac{p(i_{m+1}|i_1, \ldots, i_m, j_1, \ldots, j_l)}{p(i_{m+1}|i_1, \ldots, i_m)}, \tag{8}$$

where the parameters m and l indicate the block-lengths (=number of included past observations) in the I and J series, respectively. The sum must be taken over all possible states $i, j \in \{1, \ldots, S\}$.

It would generally be desirable to choose the parameter m as large as possible in order to find an invariant value defined as in the case of the conditional entropies for $m \to \infty$, but in practice the finite size of any real dataset imposes the need to find a reasonable compromise between unwanted finite sample effects (the amount of data required grows like $S^{(m+l)}$) and a higher closeness to the limit. A potential pitfall consists in choosing the parameter m too small, in which case information contained in past observations of actually both series may be misinterpreted as an information flow from series J to I. In a conservative approach it would thus be advisable to choose m as large as possible and set $l = 1$, which we will do in all forthcoming analyses. From equations (8) and (4) one deduces for the range of transfer entropy: $0 \le T_{J\to I}(m,l) \le H_I$.

4 Empirical results

4.1 Partitioning the data

The first step in any analysis of real data with tools based on symbolic dynamics like transfer entropy is to discretise the data by some coarse graining. Although the financial data is actually already in a discrete form, its resolution is by far too high with respect to the amount of records available. For more robust statistics and especially in the case of multi-fractal phenomena it is often recommendable to work with coverings and use generalised Renyi entropies instead of partitions and Shannon entropies [19].

In the present case, however, a straightforward implementation defining a partition with marginal equal-probability for every symbol will lead to sensible results. Such a partition is generated by dividing the range of the given dataset into S (size of the alphabet) disjoint intervals, such that the number of data points in every interval is constant and therefore $p(i) = 1/S$ and consequently $H_I = -\sum_{i=1}^{S} p(i) \log_2 p(i) = -S\frac{1}{S} \log_2 \frac{1}{S} = \log_2 S$ both automatically hold for every encoded series I, where every data point has now been uniquely replaced by the

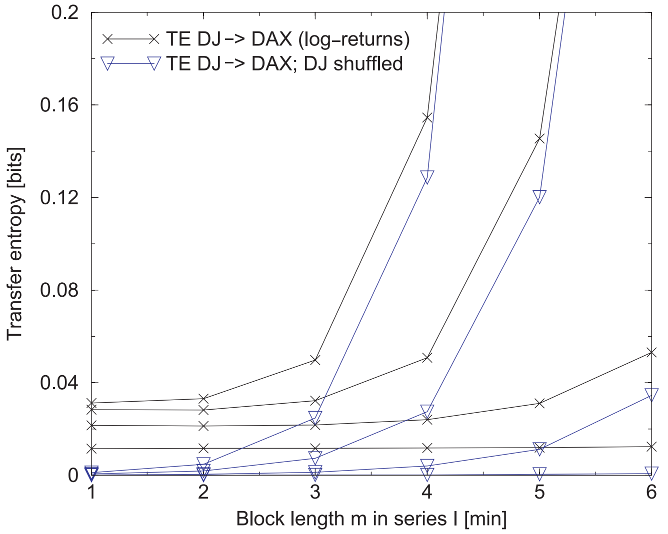

Fig. 2. Transfer entropy measuring the information flow from Dow Jones to DAX series, using various partitions of $S = 2$, 3, 4, 5 symbols (bottom to top). Upper lines have been calculated on the log-returns of DJ and DAX, for the lower ones (triangles) the log-returns of the DJ series have previously been shuffled.

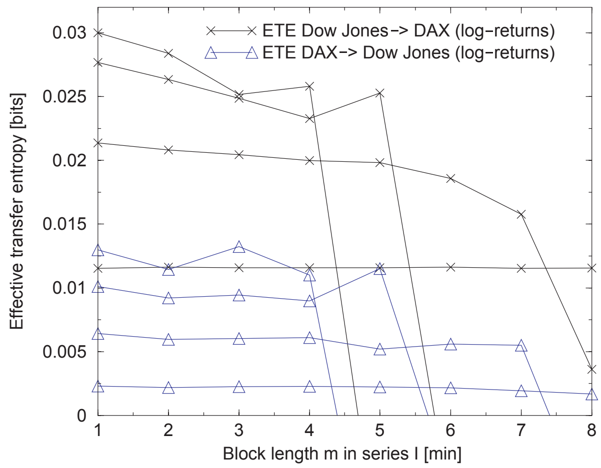

Fig. 3. Effective transfer entropy measuring the information flow between Dow Jones and DAX series, and *vice versa*, using four different partitions of $S = 2$, 3, 4, 5 symbols (bottom to top).

label of its proper interval. Apart from its simpleness, this approach has the advantage of neutralising undesirable effects due to very inhomogeneous histograms, and it also ignores the trivial information gain obtained by just observing marginal distributions. Furthermore, for data with an approximately symmetric distribution the concrete meaning of partitions consisting of few symbols is quite intuitive: two symbols ($S = 2$) only take the sign of the increments into account, three correspond to the three possible moves (i) larger gain, (ii) roughly neutral, (iii) larger loss etc. Transfer entropy will of course depend on the specific partition chosen; however, by varying the partitions one tries to find approximately invariant results.

4.2 Effective transfer entropy

In Figure 2 are displayed for four different partitions first results for the information transport from the DJ to the DAX series[6]. The steady rise of the observed transfer entropy with increasing block length m is not compatible with the theoretical expectations, and therefore no information flow can be attributed to these "raw" findings. In order to investigate their significance we now confront the obtained set of curves with a second set, calculated in exactly the same way as the first, except that the data points of the second series, that represents the source of the presumed information flow, have been shuffled, a technique related to the concept of surrogate data. With such a preprocessing, all potential correlations between the two series I and J are destroyed, and hence the observed transfer entropy should be zero, but as can be noted in Figure 2, also the new curves calculated with the shuffled DJ log-returns rise monotonically and have similar values as their original counterparts. Since there is no structure in the data, the observed non-zero values must be an artefact of the finite sample size, which also naturally accounts for the unexpected increase of the transfer entropy for growing block lengths m.

Consequently, in order for transfer entropy to objectively confirm an information flow, the empirical curves need to be above the ones generated by the shuffled data, which can be interpreted as significance threshold. At this point it is convenient to introduce a new and so far only heuristically motivated variable that facilitates the interpretation of the results: we define *effective transfer entropy* (ETE) as the difference of the usual transfer entropy calculated for the empirical series and the transfer entropy calculated for the same series, but with the J series shuffled:

$$ET_{J\rightarrow I}(m,l) := T_{J\rightarrow I}(m,l) - T_{J_{\text{shuffled}}\rightarrow I}(m,l). \quad (9)$$

In Section 5 we will demonstrate the usefulness and formal consistency of effective transfer entropy, thereby justifying a posteriori its application in the following. However, the misestimation of Shannon entropies due to finite sample effects is a well known phenomenon, see, *e.g.* [20,21].

In Figure 3 we show results for the effective transfer entropy in the case Dow Jones and DAX, considering both directions. From the now much clearer overall picture the following conclusions can be deduced:

- A flow of information from minute t of one series to the following minute of the other series is confirmed for both directions, thereby suggesting interactions between the two financial markets at a time scale of one minute or less.
- The two series do not have the same relative "weight", *i.e.* more information is transfered from the DJ to the

[6] As mentioned before, the parameter l, referring to the block-length in the J series (here DJ) has been fixed at $l = 1$.

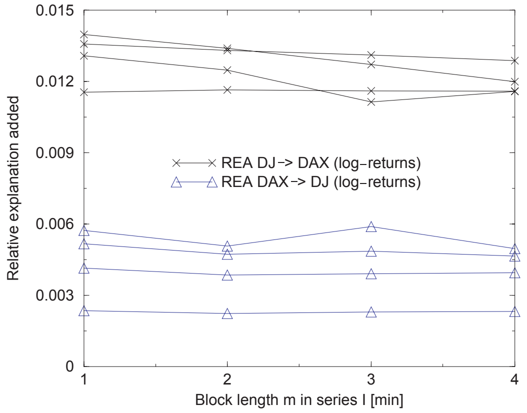

Fig. 4. Relative explanation added by the current value of the DJ for next minute's value of the DAX and *vice versa*, for four partitions of $S = 2, 3, 4, 5$ and $S = 2, 5, 3, 4$ symbols (bottom to top along y-axis at $m = 1$).

DAX than *vice versa*, which may seem trivial as a purely economical fact, but actually confirms in an independent way the validity of the transfer entropy formalism.

– The severeness of finite sample effects (how fast do the curves drop to zero for growing block-lengths m) highly depends on the size S of the alphabet. It becomes clear that the use of more than a few symbols is not compatible with the amount of data at disposal.

These conclusions have been drawn by interpreting Figure 3 qualitatively, but what about the meaning of the actual numbers? In the following we will suggest two possible interpretations.

4.3 Relative explanation added

The first and rather naive approach consists in relating the measured amount of information flow from J to I to the total flow of information in I, or, in other words, in asking of how much of $I(t)$ is additionally explained, when we already know the past of series I and then take into account the last observation of J, $J(t-1)$. Expressing this *relative explanation added* (REA) formally:

$$REA(m,l) := \frac{ET_{J\to I}(m,l)}{h_I(m)}, \qquad (10)$$

for which we report quantitative results in Figure 4[7]. Here, the finite sample effects on $h_I(m)$ are sufficiently small to be ignored. As can be noted, all curves are very robust against changes of m, the memory of the data set I. It shows that the information gained from the second time series cannot be compensated by taking into account a longer memory of the time series to be predicted. As a summary of the figure, we report the averaged values of the relative explanation added in Table 1, where the errors represent the standard deviations.

[7] It should be pointed out that the total explanatory power of $J(t-1)$ with respect to $I(t)$, *i.e.* when shared information with past observations of I is not excluded, might be much higher.

Table 1. Averaged relative explanation added: how much (here in percent) of $I(t)$ can be explained only by $J(t-1)$?

Symbols S	REA DJ → DAX [%]	REA DAX → DJ [%]
2	1.16 ± 0.04	0.23 ± 0.01
3	1.32 ± 0.03	0.40 ± 0.01
4	1.30 ± 0.09	0.49 ± 0.02
5	1.21 ± 0.09	0.54 ± 0.05

It is interesting to note the rather large gap in the values of Table 1 between $S = 2$ and $S = 3$: For the case Dow Jones → DAX all values for partitions finer than the bipartion are compatible within standard deviation. Also in the second case we observe a gap between the value found for the bipartition and all others. Since the bipartition chosen by us has the special characteristic that it can only represent a linear statistical dependence, the observed jump in the information flow when going to higher resolutions possibly implies a nonlinear correlation between the two series.

4.4 Comparison with linear autoregressive process

In a second approach aimed to interpret the quantitative results found in Section 4.2 we will propose a model that in a simplified way mimics the behaviour of the two financial time series and yet admits a clear identification of what concrete numerical values of the transfer entropy mean. Since we are investigating multivariate properties, we assume the single series to be well represented by uncorrelated Gaussian noise (r, s) with zero mean and unit standard deviation[8]. For simplicity we will consider a linear autoregressive coupling only in one direction of the following form:

$$x(t) := r(t) + \epsilon y(t-1) \quad \text{and} \quad y(t) := s(t). \qquad (11)$$

The parameter ϵ regulates the strength of the coupling and, once the number of symbols S for the partition has been fixed, also in a unique way the information flow from $y(t)$ to $x(t)$. In fact, for every set of parameters (S, ϵ) the corresponding transfer entropy can be calculated by numerically evaluating Gaussian integrals over appropriate regions in the real plane. Results of that calculation together with the corresponding values of the DJ/DAX case taken from Figure 3 (m=1) are shown in Figure 5. Two important implications follow:

– The apparently low values found for the information flow as reported in Figure 3 correspond to surprisingly strong couplings of $\epsilon \simeq 0.1$ (DAX→DJ) and $\epsilon \simeq 0.2$ (DJ→DAX). Let us note that in the case of a binary encoding $\epsilon = 0.2$ implies a correct forecast rate of 56,5%.

[8] The standard deviation actually has no influence in the case of linear coupling.

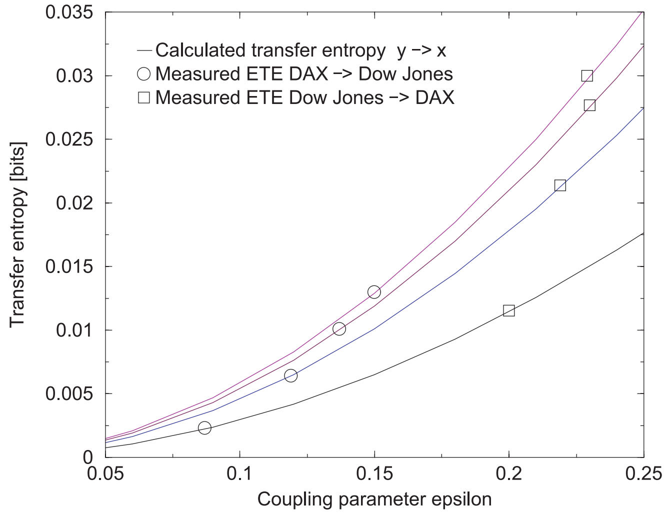

Fig. 5. Analytically derived transfer entropy as function of the coupling strength for the information flow from $y(t)$ to $x(t)$ as defined in equation (11), for $S = 2, 3, 4, 5$ symbols (from below). Indicated are also the intersections with the corresponding values for the DJ/DAX case taken from Figure 3.

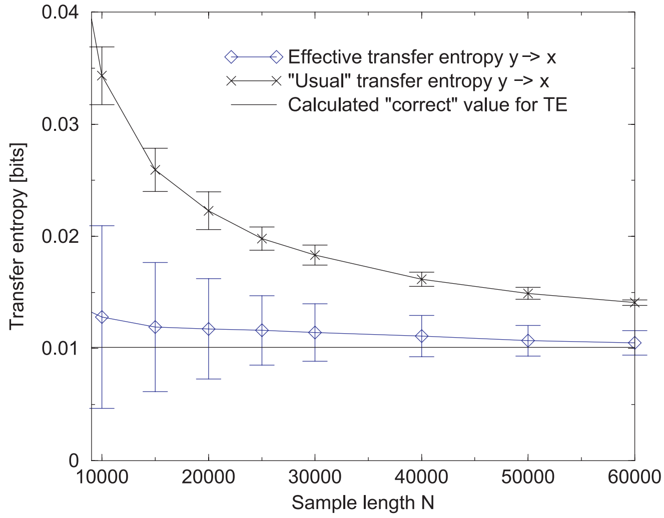

Fig. 6. Comparison of the behaviour of transfer entropy and effective transfer entropy for a varying sample size N: the information flow $y(t)$ to $x(t)$ (Eq. (11), with $\epsilon = 0.15$, $S = 3$ and $m = 4$) was measured for ten different realizations of the process, then average and standard deviation were calculated.

– The steady increase – in all except one case – of the associated coupling parameter ϵ for finer partitions strongly indicates a nonlinear coupling between the two financial series: if our linear model had been a faithful representation of the real coupling between DJ/DAX, we should have seen a clustering of the empirical ϵ values around one "correct" value. Instead we observe that the realized flow of information grows faster for finer partitions than it should be in the case of a linear coupling, meaning that the finer partitions allow to reveal some additional coupling.

5 Validity of effective transfer entropy estimates

5.1 Properties of the estimator

We now want to justify the previous use of effective transfer entropy by providing evidence for its formal consistency and improved ability to cope with finite sample effects. From its definition equation (9) one verifies immediately the correct asymptotic behaviour when the sample size N approaches infinity: $N \to \infty \Rightarrow T_{J_{\text{shuffled}} \to I}(m,l) \to 0$ by the definition of transfer entropy, which is thus retrieved.

For illustrating the improved estimation in presence of finite sample effects we will make use of the stochastic model defined earlier in equation (11). By reducing successively the length N of a time series that was generated according to the model, the response to finite sample effects of both effective and the usual transfer entropy can be simulated in a generic and – since we know the correct value from calculations – controllable way.

As can be seen from the results of that test, displayed for a typical parameter configuration in Figure 6, the effective transfer entropy indeed approaches the correct value notably faster than the usual transfer entropy estimator; moreover, though it suffers from higher fluctuations due to the summation of two random terms, it also has a smaller bias at all considered sample lengths. Returning to the main point of this section, we can conclude that the preferred application of effective transfer entropy turned out to be well justified.

5.2 Numerical and systematic errors

In the discussion of possible sources of errors, there are two aspects we retain important. The first one, concerning the stationarity of the data, constitutes a critical issue not only for this work, but for the whole data analysis related branch of econophysics. That financial data cannot be considered strictly stationary is widely accepted, but few attempts[9] have been made in order to develop statistical methods taking that into account appropriately. With reference to our case this means that we cannot assume total time independence for the single $p(i)$ and conditional $p(i|j)$ probabilities, and, in fact, in a moving window analysis fluctuations became apparent in the information flow between Dow Jones and DAX. This somewhat weakens the numerical results presented here, but the qualitative results, *i.e.* the existence of the information flow, should not be affected. Actually, the non-stationarity must not necessarily be disadvantageous, but instead could be used to identify periods of stronger and weaker coupling between the indices – of course only for large enough datasets.

Since the measurement errors in the electronically elsewhere recorded data cannot be assessed here, the remaining cause of errors in our work is given by the statistical

[9] The DFA (detrended fluctuation analysis [22]) represents one of them.

fluctuations in the performed calculations and estimates. An idea of the typical statistical error of effective transfer entropy can be read off from Figure 6. For the sample length N analysed here ($\simeq 6 \times 10^5$) the error is rather small, and is judged to be negligible in comparison to the larger fluctuations induced by the weak stationarity of the data.

6 Conclusion

Summary

In this paper, a new estimator for Schreiber's transfer entropy, called *effective transfer entropy*, was presented. By means of a significance threshold, this new estimator takes finite sample effects explicitly into account, which lead to greatly improved numerical results in the case of an empirical analysis of financial time series. In fact, profiting from the model-free approach of transfer entropy, we were able to investigate the coupling between two financial time series, namely the Dow Jones and DAX stock index. First, a significant information transfer between them could be detected, which suggests possible interactions between these financial markets at time scales of less than one minute. This result is surprising, since it suggests that a certain number of agents is trading on both markets simultaneously. However, since we are discussing the log-returns and hence the fluctuations of the stock indices, one cannot expect any information flow on larger time horizons. As a second result, the well known higher relative impact of the US-index on the German DAX has been retrieved from the data, thereby confirming the applied method.

In order to understand the numerical results, which carry the rather abstract dimension of bits, in a more intuitive way, we defined two possible interpretive instruments: first *relative explanation added*, which expresses the percentage of information contained in future observation $I(t+1)$ that can be explained only by observation $J(t)$. Here we found values slightly above one percent for what the Dow Jones explains of the DAX, and about half percent for the opposite case, again in qualitative accordance with generally known facts.

The second interpretation could be given within the context of a simple two-dimensional autoregressive process with a one-parameter adjustable linear coupling, for which the transfer entropy could be derived analytically. The following identification of the empirically observed amount of information flow with the coupling strength parameter of the model indicated a surprisingly strong coupling between the two stock indices. In addition, the observed monotonous increase of the associated coupling strengths for partitions with higher resolution showed the incompatibility of the model with the real process, implying thus nonlinearity for the coupling between Dow Jones and DAX.

Perspectives

Apart from developing forecast algorithms that exploit the identified redundancies, a possible next step following the presented work could be to set up other bivariate models with more realistic couplings and then test them by comparing their transfer entropy behaviour with the empirically observed one. A further interesting perspective consists in measuring information flows between several financial time series, *e.g.* various FX-series, thereby deriving a currency taxonomy and a hierarchy of relative "weights".

We would like to thank Deutsche Bank Research and in particular C. Kreuter for having provided the financial data.

References

1. T. Schreiber, Phys. Rev. Lett. **85**, 461 (2000)
2. J.-P. Bouchaud, M. Potters, *Theory of financial risks* (Cambridge University Press, Cambridge, 2000)
3. S. Ghashghaie, W. Breymann, J. Peinke, P. Talkner, Y. Dodge, Nature **381**, 767 (1996)
4. L. Laloux, P. Cizeau, J.-P. Bouchaud, M. Potters, Phys. Rev. Lett. **83**, 1467 (1999)
5. T. Lux, M. Marchesi, Nature **397**, 498 (1999)
6. R.N. Mantegna, H.E. Stanley, *An introduction to econophysics. Correlations and complexity in finance* (Cambridge University Press, Cambridge, 2000)
7. V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral, H.E. Stanley, Phys. Rev. Lett. **83**, 1471 (1999)
8. D. Chowdhury, D. Stauffer, Eur. Phys. J. B **8**, 477 (1999)
9. D. Sornette, A. Johansen, J.-P. Bouchaud, J. Phys. I **6**, 167 (1996)
10. R. Baviera, D. Vergni, A. Vulpiani, Physica A **280**, 566 (2000)
11. G.A. Darbellay, D. Wuertz, Physica A **287**, 429 (2000)
12. P. Gopikrishnan, M. Meyer, L.A.N. Amaral, H.E. Stanley, Eur. Phys. J. B **3**, 139 (1998)
13. L. Molgedey, W. Ebeling, Physica A **287**, 420 (2000)
14. N. Vanderwalle, M. Ausloos, Int. J. Mod. Phys. C **9**, 711 (1998)
15. S. Drożdż, F. Grümmer, F. Ruf, J. Speth, Physica A **294**, 226 (2001)
16. C.E. Shannon, W. Weaver, *The mathematical theory of information* (University of Illinois Press, Urbana IL, 1949)
17. A.N. Kolmogorov, *Information theory and the theory of algorithms*, Vol. 3 (Kluwer, Dordrecht, 1993), selected works
18. P. Billingsley, *Ergodic Theory and Information* (Wiley, New York, 1965)
19. H. Kantz, T. Schreiber, *Nonlinear time series analysis* (Cambridge University Press, Cambridge, 1997)
20. P. Grassberger, Phys. Lett. A **128**, 369 (1988)
21. H. Kantz, T. Schürmann, Chaos **6**, 167 (1996)
22. Y.H. Liu, P. Cizeau, M. Meyer, C.K. Peng, H.E. Stanley, Physica A **245**, 437 (1997)