Normalizing the causality between time series

X. San Liang[*]

*Nanjing University of Information Science and Technology (Nanjing Institute of Meteorology), Nanjing 210044,
and China Institute for Advanced Study, Central University of Finance and Economics, Beijing 100081, China*

Recently, a rigorous yet concise formula was derived to evaluate information flow, and hence the causality in a quantitative sense, between time series. To assess the importance of a resulting causality, it needs to be normalized. The normalization is achieved through distinguishing a Lyapunov exponent-like, one-dimensional phase-space stretching rate and a noise-to-signal ratio from the rate of information flow in the balance of the marginal entropy evolution of the flow recipient. It is verified with autoregressive models and applied to a real financial analysis problem. An unusually strong one-way causality is identified from IBM (International Business Machines Corporation) to GE (General Electric Company) in their early era, revealing to us an old story, which has almost faded into oblivion, about "Seven Dwarfs" competing with a giant for the mainframe computer market.

I. INTRODUCTION

Information flow, or information transfer as it may be referred to in the literature, has long been recognized as the logically sound measure of causality between dynamical events [1]. It possesses the needed asymmetry or directionalism for a cause-effect relation and, moreover, provides a quantitative characterization of the otherwise statistical test, e.g., the Granger causality test [2]. For this reason, the past decades have seen a surge of interest in this arena of research. Measures of information flow proposed thus far include, for example, time-delayed mutual information [3], transfer entropy [4], momentary information transfer [6], and causation entropy [7], among which transfer entropy has been proved to be equivalent to Granger causality up to a factor of 2 for linear systems [5].

Recently, it has been shown that the notion of information flow actually can be put on a rigorous footing within the framework of dynamical systems. The rate of information flowing from one component to another can be derived from first principles, and the expected property of causality turns out to be a proved theorem [10].

In the case where only a pair of time series, rather than the system, is given, in principle the information flow can be estimated. Particularly, under the assumption of a linear model with additive noise, the maximum likelihood estimate (MLE) of the information flow in (4) turns out to be very tight in form, involving only the common statistics, namely, sample covariances [11]. Take two series X_1 and X_2, for example. The MLE of the rate of information flowing from X_2 to X_1 is shown to be

$$T_{2\to1} = \frac{C_{11}C_{12}C_{2,d1} - C_{12}^2 C_{1,d1}}{C_{11}^2 C_{22} - C_{11}C_{12}^2}, \quad (1)$$

where C_{ij} is the sample covariance between X_i and X_j, and $C_{i,dj}$ is that between X_i and $\dot{X}_j$, $\dot{X}_j$ being the dif-

ference approximation of dX_j/dt using the *Euler forward scheme* [11]. Ideally, when $T_{2\to1} = 0$, X_2 is not the cause of X_1, and vice versa. It is easy to see that, if $C_{12} = 0$, then $T_{2\to1} = 0$, but when $T_{2\to1} = 0$, C_{12} does not need to vanish. That is, contrapositively, *causation implies correlation, but correlation does not imply causation*. (Throughout the text "causation" and "causality" are used synonymously.) In an explicitly quantitative way, this resolves the long-standing debate over causation versus correlation.

The magnitude of an information flow may differ from case to case. It needs to be normalized, just as covariance does, in order to have its importance assessed. In the extreme cases where no causality exists, though theoretically the corresponding information flow rates should be 0, in reality their estimators from time series generally do not precisely vanish. One then cannot tell whether the causality indeed exists by the absolute magnitude.

A simple example may help illustrate the issue better. Consider the series generated from two autoregressive processes,

$$X_1(n+1) = 0.1 + 0.5X_1(n) + \alpha X_2(n) + e_1(n+1), \quad (2a)$$

$$X_2(n+1) = 0.7 + \beta X_1(n) + 0.6X_2(n) + e_2(n+1), \quad (2b)$$

where the errors $e_1 \sim N(0,1)$ and $e_2 \sim N(0,1)$ are independent. Generate a pair of series with 80 000 values each[1], and perform the causality analysis. We list in Table I the

---

[*]sanliang@courant.nyu.edu

---

[1]Here the series are generated with MATLAB version 6.5.0.180913a (R13) for Windows XP version 5.1. Note that because of the pseudorandom number generator, which may not be as satisfactory as we had thought, the series thus generated may yield different information flow rates, although the mean is expected to converge to the same value if an ensemble of series is examined (thanks are due to Dionissios Hristopulos and Adolf Stips for pointing this out). The same data-generating problem exists with the touchstone stochastic differential equation in the previous publication [11]. A correctly generated series should yield a correct stationary covariance matrix, which can be computed accurately by solving a deterministic system. For easy reference, the data we used in [11] can be downloaded from http://www.ncoads.org/datasets/PRE2014.dat.gz.

TABLE I. Rates of information flow for the series generated with (2) and their respective confidence intervals at a 90% significance level. Units are 10^{-4} nats per iteration.

Case	α	β	$T_{2\to1}$	$T_{1\to2}$
I	0.5	0	1481 ± 15	-2 ± 15
II	0	0	-0.137 ± 0.219	0.128 ± 0.219
III	0.01	0.01	1.648 ± 0.692	1.123 ± 0.639

information flow rates and their respective confidence intervals at a 90% significance level. (The results may differ slightly for different series due to the pseudorandom number generation.)

For case I, $|T_{2\to1}/T_{1\to2}| > 740$, one may then conclude that this is a one-way causality from X_2 to X_1, as is indeed true. For case II, however, one actually cannot say much from the numbers. Though small, they tell no more than that the information flows in both directions are of equal importance. Of course, one may argue that the statistical significance test says it: at a 90% level these flow rates are not significantly different from 0. However, such a test just tells how precise the estimate is with the available data; it depends, for example, on the length of the series, which is irrelevant to the parameter to be estimated. In other words, an insignificant estimated rate may appear significant if more data are included. To see this more clearly, look at case III. Obviously, the information flows, albeit existent, make only tiny contributions to their respective series, as the coupling coefficients are over an order smaller; in classical perturbation analysis, they can be dropped to the first-order approximation. The computed information flows are significantly different from 0 at a 90% level: from one viewpoint, this testifies to the success of the formalism. However, the small numbers cannot tell how important they are, since, with a slowly varying series, even the dominant flow rate could be very low. On the other hand, if we cut the series by half and pick the first 40 000 points for analysis, then the results will be $T_{2\to1} = 0.653 \pm 0.751$ and $T_{1\to2} = 1.240 \pm 0.690$ (in 10^{-4} nats/iteration). So $T_{2\to1}$ is insignificant while $T_{1\to2}$ is significant. (Again, these small numbers may fluctuate due to the pseudorandom number generator.) Can one thus conclude that there is a one-way causality? or Can one thus assert that this shortened series yields a more reliable estimation? Surely this is absurd. The problem here is that we do need a normalized flow to evaluate its importance relative to other factors. In this study, we present a way to arrive at such a flow, and apply it to the analysis of several realistic financial time series.

II. INFORMATION FLOW NORMALIZATION

The normalization is not as simple as it seems to be. A natural normalizer that comes to mind, at the hint of correlation coefficient, might be the information of a series transferred from itself. A snag is, however, that this quantity may turn out to be 0, just as that in the Hénon map, a benchmark problem we have examined before (see the references in [9]). Another snag is that the above $T_{2\to1}$ and $T_{1\to2}$ usually do not share the same normalizer as that in a correlation analysis based on the Cauchy-Schwarz inequality. That is, two information flows of

equal magnitude may be of different relative importance in their respective series.

To arrive at a logically and physically sound normalizer, we need to get to the basics and analyze how an information flow within a system is derived. Consider a two-dimensional (2D) stochastic system

$$d\mathbf{X} = \mathbf{F}(\mathbf{X},t)dt + \mathbf{B}(\mathbf{X},t)d\mathbf{W}, \quad (3)$$

where $\mathbf{F} = (F_1, F_2)^T$ is the vector of drift coefficients (differentiable vector field), $\mathbf{B} = (b_{ij})$ the matrix of stochastic perturbation coefficients, and $\mathbf{W}$ a 2D standard Wiener process. Let $g_{ij} = \sum_k b_{ik}b_{jk}$ and ρ_i be the marginal probability density function of x_i. It is proved [10] that the time rate of information flowing from X_2 to X_1 is

$$T_{2\to1} = -E\left(\frac{1}{\rho_1}\frac{\partial F_1\rho_1}{\partial x_1}\right) + \frac{1}{2}E\left(\frac{1}{\rho_1}\frac{\partial^2 g_{11}\rho_1}{\partial x_1^2}\right), \quad (4)$$

where E signifies the operator mathematical expectation. This measure of information flow is asymmetric between the two parties, and particularly, if the process underlying X_1 does not depend on X_2, then the resulting information flow from X_2 to X_1 vanishes, i.e., X_2 is not causal to X_1. This is the so-called *property of causality*, a fact rigorously proven rather than just verified in applications.

When $T_{2\to1}$ is nonzero, it may take positive or negative values. Ideally, a positive $T_{2\to1}$ means that X_2 causes X_1 to be more uncertain, while a negative $T_{2\to1}$ reduces the entropy of X_1. For more details, the reader is referred to Ref. [11].

By Ref. [10], the rate of change of the marginal entropy of X_1 is

$$\frac{dH_1}{dt} = -E\left(F_1\frac{\partial \log \rho_1}{\partial x_1}\right) - \frac{1}{2}E\left(g_{11}\frac{\partial^2 \log \rho_1}{\partial x_1^2}\right). \quad (5)$$

It is actually a result of two mutually exclusive mechanisms: the first is the information flow $T_{2\to1}$ as shown in (4); the second is the complement, i.e., the rate of entropy increase without taking into account the effect of X_2. Denoting the latter $\frac{dH_{1\setminus2}}{dt}$, it has been proven in [10] that

$$\frac{dH_{1\setminus2}}{dt} = E\left(\frac{\partial F_1}{\partial x_1}\right) - \frac{1}{2}E\left(g_{11}\frac{\partial^2 \log \rho_1}{\partial x_1^2}\right)$$
$$- \frac{1}{2}E\left(\frac{1}{\rho_1}\frac{\partial^2 g_{11}\rho_1}{\partial x_1^2}\right). \quad (6)$$

The right-hand side has three terms. The first term is precisely the time rate of change of H_1 due to X_1 itself in the absence of stochasticity. This is the starting point which we showed in 2005 [8] in establishing the rigorous formalism and proved later (cf. [9]). Hence through a careful analysis, the increase in the marginal entropy H_1 is decomposed into three parts,

$$\frac{dH_1^*}{dt} = E\left(\frac{\partial F_1}{\partial x_1}\right), \quad (7)$$

$$\frac{dH_1^{\text{noise}}}{dt} = -\frac{1}{2}E\left(g_{11}\frac{\partial^2 \log \rho_1}{\partial x_1^2}\right) - \frac{1}{2}E\left(\frac{1}{\rho_1}\frac{\partial^2 g_{11}\rho_1}{\partial x_1^2}\right), \quad (8)$$

and $T_{2\to1}$ in (4), which correspond to, respectively, the phase-space expansion along the X_1 direction, the stochastic effect, and the information flowing from X_2; Fig. 1 shows a schematic.
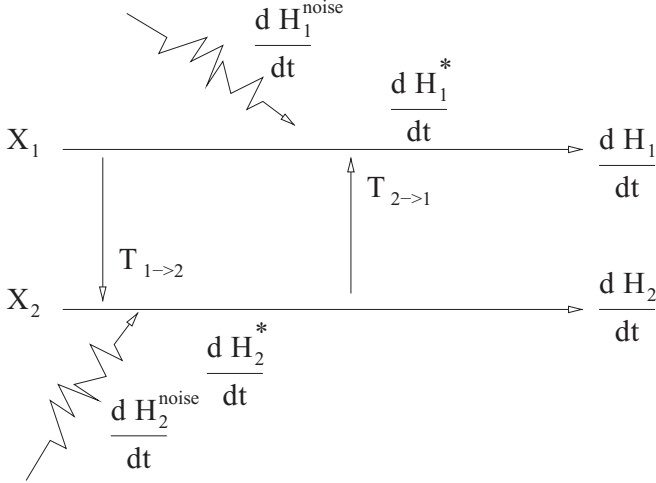
FIG. 1. Schematic of the marginal entropy evolutions and information flows in the system of (X_1, X_2).

Note that this decomposition does not appear explicitly in the marginal entropy evolution equation, (5), as the two stochastic terms cancel out.

The normalization is now made easy. Let

$$Z_{2\to1} \equiv |T_{2\to1}| + \left|\frac{dH_1^*}{dt}\right| + \left|\frac{dH_1^{\text{noise}}}{dt}\right|. \tag{9}$$

Obviously it is no less than $T_{2\to1}$ in magnitude and cannot be 0 unless X_1 does not change, a situation that is excluded in time series analysis. We may therefore pick $Z_{2\to1}$ as the normalizer and define

$$\tau_{2\to1} = T_{2\to1}/Z_{2\to1}. \tag{10}$$

This way if $\tau_{2\to1} = 1$, the variation of H_1 is 100% due to the information flow from X_2; if $\tau_{2\to1}$ is approximately 0, X_2 is not the cause. Therefore, $\tau_{2\to1}$ assesses the importance of the influence of X_2 on X_1 relative to other processes.

It should be pointed out that the above normalizer applies to $T_{2\to1}$ only. For $T_{1\to2}$, it is

$$Z_{1\to2} = |T_{1\to2}| + \left|\frac{dH_2^*}{dt}\right| + \left|\frac{dH_2^{\text{noise}}}{dt}\right|,$$

which may be quite different in value. This, from another viewpoint, reflects the asymmetry between $T_{2\to1}$ and $T_{1\to2}$.

III. ESTIMATION

As in Ref. [11], consider a linear version of the stochastic differential equation, (3),

$$d\mathbf{X} = (\mathbf{f} + \mathbf{A}\mathbf{X})dt + \mathbf{B}d\mathbf{W}, \tag{11}$$

where $\mathbf{f}$ is a constant vector, and $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$ are constant matrices. Initially if $\mathbf{X}$ obeys a normal distribution, then it is normally distributed forever. Let the mean and covariance matrix be $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma} = (\sigma_{ij})$. They evolve according to

$$\frac{d\boldsymbol{\mu}}{dt} = \mathbf{f} + \mathbf{A}\boldsymbol{\mu}, \tag{12}$$

$$\frac{d\boldsymbol{\Sigma}}{dt} = \mathbf{A}\boldsymbol{\Sigma} + \boldsymbol{\Sigma}\mathbf{A}^T + \mathbf{B}\mathbf{B}^T. \tag{13}$$

So Eqs. (7) and (8) can be explicitly evaluated:

$$\frac{dH_1^*}{dt} = E(a_{11}) = a_{11}, \tag{14}$$

and

$$\begin{aligned}
\frac{dH_1^{\text{noise}}}{dt} &= -\frac{1}{2}E\left(g_{11}\frac{\partial^2 \log \rho_1}{\partial x_1^2}\right) - \frac{1}{2}E\left(\frac{1}{\rho_1}\frac{\partial^2 g_{11}\rho_1}{\partial x_1^2}\right) \\
&= -\frac{1}{2}E\left[g_{11}\left(-\frac{1}{\sigma_{11}}\right)\right] - \frac{1}{2}\iint_{\mathbb{R}^2}\rho_{2|1}\frac{\partial^2 g_{11}\rho_1}{\partial x_1^2}dx_1 dx_2 \\
&= \frac{1}{2}\frac{g_{11}}{\sigma_{11}} - \frac{1}{2}\int_{\mathbb{R}}\frac{\partial^2 g_{11}\rho_1}{\partial x_1^2}\left(\int_{\mathbb{R}}\rho_{2|1}(x_2|x_1)dx_2\right)dx_1,
\end{aligned}$$

since neither g_{11} nor ρ_1 depends on x_2. But $\int_{\mathbb{R}}\rho_{2|1}dx_2 = 1$, and ρ_1 is compactly supported, so the whole second term on the right-hand side then vanishes. Hence

$$\frac{dH_1^{\text{noise}}}{dt} = \frac{1}{2}\frac{g_{11}}{\sigma_{11}}. \tag{15}$$

Equations (14) and (15), together with the information flow from X_2 to X_1 as we have obtained before [8,11], $T_{2\to1} = \frac{\sigma_{12}}{\sigma_{11}}a_{12}$, form the three constituents that account for the evolution of the marginal entropy of X_1.

An observation about $\frac{d}{dt}H_1^{\text{noise}} = g_{11}/(2\sigma_{11})$, where $g_{11} = b_{11}^2 + b_{12}^2$, is that it is always positive. That is, noises always contribute to an increase in the marginal entropy of X_1, conforming to common sense. In financial economics, this reflects the volatility of, say, a stock. On the other hand, for a stationary series, the left-hand side of Eq. (13) tends to be 0, and the balance on the right-hand side requires that $2\sigma_{11} \sim g_{11}$. So this quantity is also related to the noise-to-signal ratio.

The above results need to be estimated if what we are given is just a pair of time series. That is, what we know is a single realization of some unknown system, which, if known, can produce infinitely many realizations. The problem now becomes estimating (14) and (15) with the available statistics of the given time series.

We use maximum likelihood estimation to achieve the goal. The procedure follows precisely that in [11], to which we refer the reader for details. Suppose that the series are sampled at regular instants with a time step size Δt, and let N be the sample size. Further assume that $b_{12} = 0$ (hence $g_{11} = b_{11}^2$). We have shown that the MLEs are [11] $\hat{a}_{11} = p$, $\hat{a}_{12} = q$, $\hat{f}_1 = \dot{\bar{X}}_1 - p\bar{X}_1 - q\bar{X}_2$ (overbar signifies sample mean), with

$$p = \frac{C_{22}C_{1,d1} - C_{12}C_{2,d1}}{\det \mathbf{C}}, \tag{16}$$

$$q = \frac{-C_{12}C_{1,d1} + C_{11}C_{2,d1}}{\det \mathbf{C}}, \tag{17}$$

where $C_{i,j}$ is the sample covariance between X_i and X_j, and $C_{i,dj}$ the sample covariance between X_i and $\dot{X}_j \approx \left\{\frac{X_j(n+k)-X_j(n)}{k\Delta t}\right\}$. (Usually $k = 1$ should be used to ensure accuracy, but in some cases of deterministic chaos, where the sampling is at the highest resolution, one needs to choose $k = 2$.) The MLE of g_{11} can be obtained by computing $\hat{g}_{11} = \Delta t\overline{[\dot{X}_1 - (\hat{f}_1 + \hat{a}_{11}X_1 + \hat{a}_{12}X_2)]^2}$.

On the other hand, the population covariance matrix $\boldsymbol{\Sigma}$ can be rather accurately estimated by the sample covariance

matrix **C**. So (14) and (15) become, after some algebraic manipulation,

$$\frac{dH_1^*}{dt} = p, \tag{18}$$

$$\frac{dH_1^{\text{noise}}}{dt} = \frac{\Delta t}{2C_{11}}(C_{d1,d1} + p^2 C_{11} + q^2 C_{22} - 2pC_{d1,1} - 2qC_{d1,2} + 2pqC_{12}). \tag{19}$$

As in Ref. [11] with $T_{2\rightarrow1}$, here $\frac{dH_1^*}{dt}$ and $\frac{dH}{dt}\,_1^{\text{noise}}$ (and $Z_{2\rightarrow1}$ in the following) should bear a hat, since they are the corresponding estimators. We abuse the notation a little bit to avoid notational complexity; from now on they should be understood as their respective estimators. With these the normalizer is

$$Z_{2\rightarrow1} = |T_{2\rightarrow1}| + \left|\frac{dH_1^*}{dt}\right| + \left|\frac{dH_1^{\text{noise}}}{dt}\right|, \tag{20}$$

and hence we have the relative information flow from X_2 to X_1:

$$\tau_{2\rightarrow1} = T_{2\rightarrow1}/Z_{2\rightarrow1}. \tag{21}$$

$\tau_{1\rightarrow2}$ can be obtained simply by swapping the indices in T and Z and their respective expressions.

IV. THE AUTOREGRESSIVE EXAMPLE REVISITED

We return to the autoregressive process exemplified in the beginning. When $\alpha = 0$, $\beta = 0$, the computed relative information flow rates are

$$\tau_{2\rightarrow1} = -0.0016\%, \quad \tau_{1\rightarrow2} = 0.0018\%.$$

Clearly both are negligible in comparison to the contributions from the other processes in their respective series. For the case $\alpha = \beta = 0.01$, where one may encounter difficulty due to the ambiguous small numbers, the computed relative information flow rates are

$$\tau_{2\rightarrow1} = 0.018\%, \quad \tau_{1\rightarrow2} = 0.015\%.$$

Again, they are essentially negligible, just as one would expect.

It should be pointed out that the relative information flow, say, $\tau_{2\rightarrow1}$, makes sense only with respect to X_1, since the comparison is within the series itself. Now consider the following situation: for a two-way causal system with absolute information flows $T_{2\rightarrow1}$ and $T_{1\rightarrow2}$ of equal importance, their relative importance within their respective series could be quite different. For example,

$$X_1(n+1) = -0.5X_1(n) + 0.9X_2(n) + 2e_1(n+1),$$

$$X_2(n+1) = -0.2X_1(n) + 0.5X_2(n) + e_2(n+1),$$

e_1 and e_2 being the same as before. Initialize them with random values between [0,1] and generate 80 000 data points (on the same platform as before). The computed information flow rates (in nats per iteration), $|T_{2\rightarrow1}| = 0.13$ and $|T_{1\rightarrow2}| = 0.12$, are almost the same. The relative information flows, however, are quite different: $|\tau_{2\rightarrow1}| = 6.7\%$ and $|\tau_{1\rightarrow2}| = 13\%$.

Generally speaking, the above imbalance is a rule, not an exception, reflecting the asymmetry of information flow. One may reasonably imagine that, in some extreme situation, a flow might be dominant while its counterpart is negligible within their respective series, although the two are of the same order in absolute value.

V. APPLICATION

We now demonstrate a real-world application with several financial time series. Here it is not our intention to conduct financial economics research or study market dynamics from an econophysical point of view; our purpose is to demonstrate a brief application of the aforementioned formalism for time series analysis. Nonetheless, this topic is indeed of interest to both physicists and economists in the field of macroscopic econophysics; see, for example, [12].

We pick nine stocks in the United States and download their daily prices from YAHOO! FINANCE. These stocks are MSFT (Microsoft Corporation), AAPL (Apple Inc.), IBM (International Business Machines Corporation), INTC (Intel Corporation), GE (General Electric Company), WMT (Wal-Mart Stores Inc.), XOM (Exxon Mobil Corporation), CVS (CVS Health Corporation), and F (Ford Motor Corporation). Among these are high-tech companies (MSFT, AAPL, IBM, INTC), retail trade companies [e.g., drugstore chains (CVS) and discount stores (WMT)], the automotive industry (F), the oil and gas industry (XOM), and the multinational conglomerate corporation GE, which operates through the segments of energy, technology infrastructure, capital finance, etc. Here by "daily" we mean on a trading-day basis, excluding, say, holidays and weekends. Since stock prices are generally nonstationary, we check the series of daily return, i.e., $R(t) = [P(t + \Delta t) - P(t)]/P(t)$, or log-return, $r(t) = \ln P(t + \Delta t) - \ln P(t)$, where $P(t)$ are the adjusted closing prices in the YAHOO! spreadsheet, and Δt is 1 trading day. Following most people we use the series of log-returns r for our purpose. In fact, the return and log-return series are approximately equivalent, particularly in the high-frequency regime, as indicated in [13]. Since the most recent stock, MSFT, started on March 13, 1986, all the series are chosen from that date through December 26, 2014, when this study commenced. This amounts to 7260 data points, hence 7259 points for the log-return series.

Using Eq. (1), we compute the information flows between the nine stocks and form a matrix of flow rates; see Table II. The flow direction is represented by the matrix indices; more specifically, it is from the row index to the column index. For example, listed at location (2,4) is $T_{2\rightarrow4}$, i.e., $T_{\text{AAPL}\rightarrow\text{INTC}}$, the flow rate from Apple to Intel, while (4,2) stores the rate of the reverse flow, $T_{\text{INTC}\rightarrow\text{AAPL}}$. Also listed in the table are the respective confidence intervals at the 90% level.

In Table II, most of the information flow rates are significant at the 90% level, as highlighted. Their values vary from 4 to 22 (units, 10^{-3} nats/day; same below). The maximum is $|T_{\text{IBM}\rightarrow\text{XOM}}|$, and second to it are $|T_{\text{WMT}\rightarrow\text{CVS}}|$ and $|T_{\text{CVS}\rightarrow\text{GE}}|$.

Look at the table row by row (companies as drives). Perhaps the most conspicuous feature is that the whole CVS row is significant. Next to it is XOM, with only three insignificant entries. That is, CVS is found to be causal to all other stocks, though the causality magnitudes have yet to be assessed (see below). This does make sense. As a chain of convenience stores, CVS connects most of the general consumers and

TABLE II. Rates of absolute information flow among the nine chosen stocks (in 10^{-3} nats per trading day). For each entry the direction is from the row index to the column index of the matrix. Also listed are the standard errors at a 90% significance level (significant flows are highlighted).

	MSFT	AAPL	IBM	INTC	GE	WMT	XOM	CVS	F
MSFT	/	5 ± 7	-3 ± 8	-12 ± 11	1 ± 8	-1 ± 6	-12 ± 6	10 ± 4	3 ± 5
AAPL	-2 ± 7	/	-11 ± 7	-2 ± 9	-7 ± 6	-4 ± 5	-11 ± 4	4 ± 3	-2 ± 4
IBM	0 ± 8	5 ± 7	/	-9 ± 9	-8 ± 9	-11 ± 6	-22 ± 7	6 ± 4	1 ± 6
INTC	16 ± 11	10 ± 9	-7 ± 9	/	0 ± 8	-2 ± 6	-12 ± 5	11 ± 4	3 ± 6
GE	2 ± 8	-3 ± 6	-13 ± 9	-16 ± 8	/	-10 ± 9	-6 ± 9	14 ± 6	6 ± 9
WMT	10 ± 6	7 ± 5	4 ± 6	-5 ± 6	6 ± 9	/	0 ± 6	21 ± 7	9 ± 5
XOM	-10 ± 6	-3 ± 4	-14 ± 7	-13 ± 5	-15 ± 9	-17 ± 6	/	4 ± 5	1 ± 6
CVS	-9 ± 4	-5 ± 3	-12 ± 4	-11 ± 4	-21 ± 6	-17 ± 7	-17 ± 5	/	-7 ± 4
F	0 ± 5	0 ± 4	0 ± 6	-10 ± 6	6 ± 9	-13 ± 5	1 ± 6	6 ± 4	/

commodities and hence the corresponding industries. For XOM, it is causal because oil or gas is definitely one of the most fundamental components of the American economy. Another interesting observation is that $|T_{F \to WMT}| > |T_{F \to CVS}|$. This is easy to understand, as we rely on our motor vehicles to shop at Wal-Mart, while CVS stores could be right in the neighborhood.

The above significant absolute information flows, large or small, still need to assessed regarding their respective relative importance before any conclusion of causality is reached. Using Eq. (21), we compute the relative information flow rates (as a percentage) and list them in Table III. For clarity, those greather than or equal to 1% are highlighted. In contrast to Table II, we see only a few information flows that account for more than 1% of their respective fluctuations. This echoes what we introduced in the beginning: though significant, some information flows may be negligible in their own marginal entropy balances.

It should be noted that the causal relations generally change with time. If the series are long enough, we may look at how these information flows may vary from period to period. Pick the pair (IBM, GE) as an example. For the duration (March 1986 through present) considered above, $T_{GE \to IBM} = -13 \pm 9$, while $T_{IBM \to GE}$ is not significant. Neither $\tau_{GE \to IBM}$ nor $\tau_{IBM \to GE}$ reaches 1%. Since at the YAHOO! site both GE and IBM can be dated back to January 2, 1962, we can extend the time series a lot, up to 13 338 data points.

TABLE III. As Table II, but for relative information flow (as a percentage).

	MSFT	AAPL	IBM	INTC	GE	WMT	XOM	CVS	F
MSFT	/	0.3	-0.2	-0.8	0.0	0.0	-0.7	0.6	0.2
AAPL	-0.1	/	-0.7	-0.1	-0.5	-0.2	-0.7	0.2	-0.1
IBM	0.0	0.3	/	-0.6	-0.5	-0.7	-1.3	0.4	0.1
INTC	1.0	0.7	-0.4	/	0.0	-0.1	-0.7	0.7	0.2
GE	0.1	-0.2	-0.8	1.0	/	-0.6	-0.3	0.9	0.4
WMT	0.6	0.4	0.2	-0.3	0.4	/	0.0	1.3	0.6
XOM	-0.6	-0.2	-0.9	-0.9	-1.0	-1.1	/	0.3	-0.1
CVS	-0.6	-0.3	-0.8	-0.7	-1.3	-1.1	-1.1	/	-0.5
F	0.0	0.0	0.0	-0.6	0.4	-0.8	0.0	0.4	/

Computation of information flows with the whole series (13 338 points) results in $\tau_{IBM \to GE} = 1.6\%$, $\tau_{GE \to IBM} = -0.5\%$ and in $T_{IBM \to GE} = (27 \pm 6) \times 10^{-3}$, $T_{GE \to IBM} = (7 \pm 6) \times 10^{-3}$ nats/day, both being significant at the 90% level. This is very different from what is shown in Tables II and III, with the causal structure changed from a weak two-way causality to a stronger and more or less one-way causality. Since in the above, only data for the most recent 30 years are used, we expect that in the early years this causal structure could be much enhanced. Choosing the first 7000 points (from January 1962 through November 1989), the computed relative information flow rates are

$$\tau_{IBM \to GE} = 3.1\%, \quad \tau_{GE \to IBM} = -0.2\%;$$
$$T_{IBM \to GE} = 54 \pm 8, \quad T_{GE \to IBM} = 3 \pm 8,$$

(recall that the units for T are 10^{-3} nats/day). Narrow the period down further, to 2250–3250 (corresponding to the period 1971–1975); then

$$\tau_{IBM \to GE} = 5.7\%, \quad \tau_{GE \to IBM} = -0.99\%;$$
$$T_{IBM \to GE} = 101 \pm 21, \quad T_{GE \to IBM} = 14 \pm 21,$$

attaining a maximum of $\tau_{IBM \to GE}$, in contrast to the insignificant flow in Table II. Obviously, during this period, the causality can be approximately viewed as one-way, i.e., from IBM to GE. And the relative flow is more than 5%, much larger than the values in Table III.

The above remarkable causal structure for that particular period actually can trace its roots back to the history of GE [14]. There was such a period in the 1960s when "Seven Dwarfs" (Burroughs, Sperry Rand, Control Data, Honeywell, General Electric, RCA, and NCR) competed with IBM, the giant, for computer business, particularly, to build mainframes. In 1965, GE had only a 3.7% market share of the industry, though it was then dubbed the "King of the Dwarfs,' while IBM had a 65.3% share. Historically GE was once the largest computer user besides the U.S. Federal Government; it got into computer manufacturing to avoid dependency on others. And, indeed, throughout the 60s, the causalities between GE and IBM are not significant. Then why, as the 70s began, did the information flow from IBM to GE suddenly increase to its highest level? It turned out that GE sold its computer division to Honeywell in 1970; in the following years (starting from 1971), GE relied a

great deal on IBM products. Clearly, the GE computer history does substantiate the existence of a causation between GE and IBM and, to be more precise, an essentially one-way causation from IBM to GE. In an era when this has almost been forgotten (one cannot even find it at GE's Web site), and GE may have left the impression that it never built any computers, let alone a series of mainframes, this finding, which is solely based on the causality analysis of a couple of time series, is indeed remarkable.

VI. CONCLUDING REMARKS

To assess the importance of a flow of information from one time series, say X_2, to another, say X_1. it needs to be normalized. Getting down to the fundamentals, we were able to distinguish three types of mechanisms that contribute to the evolution of the marginal entropy of the recipient X_1—namely, the phase-space expansion in the X_1 direction, the information flow from X_2, and the contribution from noise—and hence proposed an approach to normalization. The resulting scheme is described by Eqs. (16)–(21).

It should be noted that a relative information flow is for comparison purposes within its own series. The two reverse flows between two series can only be compared in terms of absolute value, since they belong to different series. It is quite normal that two identical information flows may differ significantly in relative importance with respect to their own series, as demonstrated in our examples, reflecting the property of flow asymmetry. In some extreme situation, a pair of equal flows may have one dominant but another negligible in their respective entropy balances. In this sense, absolute and relative information flows usually need to be examined simultaneously in realistic applications.

[1] For example, K. Hlaváčková-Schindler, M. Paluš, M. Vejmelka, and J. Bhattacharya, Phys. Rep. **441**, 1 (2007); J. Pearl, *Causality: Models, Reasoning, and Inference* (MIT Press, Cambridge, MA, 2009); M. Lungarella, K. Ishiguro, Y. Kuniyoshi, and N. Otsu, Int. J. Bifurcat. Chaos **17**, 903 (2007); P. Duan, F. Yang, T. Chen, and S. L. Shah, IEEE Trans. Control Syst. Tech. **21**, 2052 (2013).

[2] C. Granger, Econometrica **37**, 424 (1969).

[3] J. A. Vastano and H. L. Swinney, Phys. Rev. Lett. **60**, 1773 (1988).

[4] T. Schreiber, Phys. Rev. Lett. **85**, 461 (2000).

[5] L. Barnett, A. B. Barrett, and A. K. Seth, Phys. Rev. Lett. **103**, 238701 (2009).

[6] J. Runge, J. Heitzig, N. Marwan, and J. Kurths, Phys. Rev. E **86**, 061121 (2012).

[7] J. Sun and E. Bolt, Physica D **267**, 49 (2014).

[8] X. S. Liang and R. Kleeman, Phys. Rev. Lett. **95**, 244101 (2005)

[9] X. S. Liang, Entropy **15**, 327 (2013).

[10] X. S. Liang, Phys. Rev. E **78**, 031113 (2008).

[11] X. S. Liang, Phys. Rev. E **90**, 052150 (2014).

[12] R. Marshinski and H. Kantz, Eur. Phys. J. B **30**, 275 (2002).

[13] R. N. Mantegna and H. E. Stanley, *An Introduction to Econophysics. Correlations and Complexity in Finance* (Cambridge University Press, Cambridge, 2000)

[14] http://en.wikipedia.org/wiki/General_Electric/.