# A Survey of Causality and Directed Information

John A.W.B. Costanzo & Jonathan Dunstan

October 6, 2014

**Abstract**

*Causality has been an important concept in philosophy at least since the days of Aristotle and his "four causes" theory. Just for how long causality has been studied mathematically is uncertain due to the often blurred lines between mathematics and philosophy, but even today assigning causality (let alone culpability) to a pair of events requires some degree of human intuition. This paper will survey Granger Causality, a tool with applications in aiding the researcher in determining causal influence in complex systems. We will also present Directed Information, a generalization of causality with applications beyond the philosophical. Discussion of the philosophical consequences of these tools is intertwined throughout, just as philosophy is intertwined with science and engineering in life.*

## 1. Introduction

Causality has been an important concept in philosophy at least since the days of Aristotle and his "four causes" theory. In essence, causality is the concept that unifies the answers to the question, "Why?". Just for how long causality has been studied mathematically is uncertain due to the often blurred lines between mathematics and philosophy, but even today assigning causality (let alone culpability) to a pair of events requires some degree of human intuition.

Determining causality algorithmically is of great interest to researchers as it would allow faster, richer analysis of complex systems in biology and economics, accelerate and robustify bug detection in software and electronics engineering, increase the effectiveness of optimal control in dynamical systems, and a host of other applications.

This paper will survey a few mathematical tools that are used to aid the researcher in determining causality between processes, how this is extended to multiple process systems, and survey a generalization of causality with applications beyond the philosophical.

Section 2 will survey the history of the statistical framework behind the quest to quantify causality, and remark on the pitfalls of using them haphazardly. The section begins with a brief summary of the most relevant notation common to most of the literature. Section 2.1 covers some of the earliest attempts to verify causality experimentally, culminating with the definition of Granger Causality. Section 2.2 begins with the definition of directed information. The properties, implications, and consequences of directed information will be discussed. Finally, section 2.3 will discuss how these methods can be used to infer direct causal links between processes in the context of complex, multiple process systems, as well as how they might imply false causal links.

Section 3 will survey a few applications of causality and directed information. A brief list of notable papers using either method will be presented. Sections 3.1 and 3.2 serve to summarize and aid the reader in the review of two studies in particular. Section 3.1 will survey "On Directed Information and Gambling" [1], showing that directed information, far from being an arbitrary, theoretical quantity, has a numeric meaning in the growth rate of a gambler's portfolio. Section 3.2 will survey "Twitter Mood Predicts the Stock Market" [2].

## 2. Definition of Granger Causality and Directed Information

We will look at processes $X^n = \{X_1, X_2, \ldots, X_n\}$, $Y^n = \{Y_1, Y_2, \ldots, Y_n\}$ as sequences of random variables. The notation $X_i^j = \{X_i, X_{i+1}, \ldots, X_j\}$.

The notation $X^n||Y^{n-d}$ is read "$X^n$ causally conditioned on $Y^{n-d}$". This notation was introduced by Kramer [3]

and refers to the random variable with distribution

$$p_{x^n||y^{n-d}} = \Pi_{i=1}^{n} p_{x_i|x^{i-1},y^{i-d}}. \quad (1)$$

This definition coincides with the definition of causally conditioned entropy, where entropy is defined as

$$H(X) = \mathbb{E}[-\log p(X)]. \quad (2)$$

Using the chain rule, the causally conditioned entropy is

$$\begin{aligned} H(X^n||Y^n) &= \mathbb{E}[-\log p(X^n||Y^n)] \\ &= \sum_{i=1}^{n} H(X_i \mid X^{i-1}, Y^i). \end{aligned}$$

## 2.1. Granger Causality

Over the centuries, scientists and philosophers alike have struggled in interpreting causality. While the statement "A causes B" appears to be simple, the specifics of what it implies are not always clear. Does B always follow A? Can A and B occur at the same time? Many attempts have been made at formalizing what causation means, and how it should be quantified. One such attempt was made by Suppes in his book *A Probabilistic Theory of Causality* , where he defined event $B_{t'}$ to be a prima facaie cause of $A_t$ if and only if

$$t' < t \quad (3)$$

$$P(B_{t'}) > 0 \quad (4)$$

$$P(A_t \mid B_{t'}) > P(A_t) \quad (5)$$

[4]

Or in other words, if B can happen, and happen before A, and A is more likely to occur when B happens, then B causes A prima facaie.

This approach is not without its flaws, however. For example, if $B_{t'}$ is the event where a woman takes birth control pills at time $t'$, and $A_t$ is the event of that she not become pregnant at time $t$, then it makes sense that taking birthcontrol prima facaie caused her to avoid becoming pregnant. But $A$ and $B$ are arbitrarily defined (within the constraints of the first two equations). So by taking the complement of both $A$ and $B$, such that $B_{t'}$ represents her forgetting to take the pill, and $A_t$ is her becoming pregnant, then from Suppes' formulation, we would conclude that not taking birth control causes pregnancy. This limitation, although seemingly pedantic, raises the widely-debated question of whether it even makes sense to discuss prima facaie causality outside the realm of tighly controlled experiments [5].

Another probabilistic approach to causality, formulated by Granger, involves examining the distributions of processes $X$ and $Y$, and how additional knowledge of one affects the distribution of the other. More formally, $\Omega_i$ represents the universal set of information. Within it is the information available to the experimenter at a given time, $Ji$, which includes past information ($t < i$). We also define $J_i$ to exclude information about $Y$, and $J_i'$ to include it. In other words,

$$J_i \subset \Omega_i \setminus Y_i \quad (6)$$

$$J_i' \subset \Omega_i \cup Y_i \quad (7)$$

With these defintions, Granger cites three main results [6]. Firstly, that if

$$F(X_{i+1} \mid J_i) = F(X_{i+1} \mid J_i') \quad (8)$$

then $Y$ does not causally influence $X$ (with respect to $J$). In other words, if information about $Y$doesn't change what we know about $X$, then $Y$ cannot be causally influencing it.

Secondly, if,

$$F(X_{i+1} \mid \Omega_i) \neq F(X_{i+1} \mid \Omega_i \setminus Y_i) \quad (9)$$

then $Y$ causally influences $X$.

The major problem with these definitions is that demonstrating causal influence requires knowledge of the entire universe. In a practical sense, this is unfeasible, since no experimenter has access to all possible variables. But given a set of information assumed to be relevent ($J$), one can determine their prima facie relation.

If,

$$F(X_{i+1} \mid J_i) \neq F(X_{i+1} \mid J_i') \tag{10}$$

Then $Y$ causes $X$ prima facie with respect to $J$. The distinction between '$Y$ causes $X$' and '$Y$ causes $X$ prima facie' is that the latter indicates that not all information is present. There could exist another variable, $Z \in \Omega_i$, which influences both $X$ and $Y$, but was neglected by experimenters.

In a practical sense, granger causality asks 'given previous values of $Y$, how well can we forecast $X$?' If $Y$ is a stochastic process, then we could try predicting $Y_i$ by performing a linear regression on its $r$ previous values [7],

$$Y_i = \sum_{j=1}^{r} a_j Y_{i-j} + \nu_i \tag{11}$$

where $a_j$'s are constants from the regression, $\nu_i$ is the error term (with variance $var(\nu_i)$), and $r$ is the number of previous values of $Y$ we wish to consider.

We could also try predicting $Y_i$ with a linear regression which includes past values of $X$.

$$\widetilde{Y}_i = \sum_{j=1}^{r} \left(a_j Y_{i-j} + b_j X_{i-j}\right) + \tilde{\nu}_i \tag{12}$$

where $\widetilde{Y}_i$ is our prediction for $Y$ including past values of $X$, and new error term $\tilde{\nu}_i$ (with variance $var(\tilde{\nu}_i)$).

To quantify the causal nature of these predictions, the logarithmic ratio of the two residual variances are examined. Specifically,

$$G_{X\to Y} = log(\frac{var(\nu_i)}{var(\tilde{\nu}_i)}) \tag{13}$$

So in other words, if considering past values of $X$ helps in forecasting $Y$, then the including $X$ will lower the residual variance $var(\tilde{\nu}_i)$, thereby increasing $G_{X\to Y}$. In this case, it is said that $X$ Granger causes $Y$ [7]. Note that these terms can be flipped around, we can just as easily ask if $Y$ causes $X$ by using the same method, and get a higher or lower $G$. This demonstrates the directionality of information flow, which will be discussed in the next section.

## 2.2. Directed Information

The concept of Directed Information first appeared in [8]. The idea was refined somewhat by Massey [9] and in its current form is defined:

$$I(X^n \to Y^n) = \sum_{i=1}^{n} I(X^i; Y_i \mid Y^{i-1}). \tag{14}$$

Contrast with the definition of mutual information of two processes:

$$I(X^n; Y^n) = \sum_{i=1}^{n} I(X^n; Y_i \mid Y^{i-1}). \tag{15}$$

Looking at the term associated with the $i^{th}$ term in the sequence, we see that the difference between mutual information and directed information is that mutual information compares $Y_i$ with the entirety of the process $X^n$, whereas directed information only compares $Y_i$ with $X^i$, the result of the process $X$ up until timestep $i$. This has the effect of capturing the "current" trend of $X$. [10] provides a simple example to show how directed information is more revealing about causal links than mutual information alone.

Let $X_i \sim Ber(\frac{1}{2})$ be $iid$ for $i = 0, 1 \ldots$ and let $Y_{i+1} = X_i$. Clearly, $X$ influences $Y$ directly, and not the other way around. Indeed,

$$\mathcal{I}(X;Y) = 1; \tag{16}$$
$$\mathcal{I}(X \to Y) = 1; \tag{17}$$
$$\mathcal{I}(Y \to X) = 0, \tag{18}$$

where $\mathcal{I}(X) = \lim_{n\to\infty} \frac{1}{n} I(X^n)$ is the mutual information rate and $\mathcal{I}(X \to Y) = \lim_{n\to\infty} \frac{1}{n} I(X^n \to Y^n)$ is the directed information rate. Notice that, even though knowing all of $Y$ is sufficient to know all of $X$, the *directed* information from $Y$ to $X$ is zero.

There is an obvious time dependence here; if we'd instead defined $Y_{i-1} = X_i$ or even $Y_{i-1} = f(X_i)$ then the results would be reversed, and it would appear that $Y$ is causing $X$ instead of the other way around—even though the phrasing of the definition suggests that $Y$ is caused by $X$. Obviously in practice it could

never happen that the current state of one process is influenced by a *future* state of another process, because the causative stimulus would have not happened yet; however, this does emphasize that some degree of time synchronicity should be employed in measuring the system.

#### 2.2.1 Conservation Law, or, Directed Information "Partitions" Mutual Information

It is known that

$$I(X^n \to Y^n) + I(0 * Y^{n-1} \to X^n) = I(X^n; Y^n). \quad (19)$$

The proof uses an inductive argument [11]. By definition, $I(X^1 \to Y^1) = I(X^1; Y^1)$ and $I(0 \to X_1) = 0$, establishing the base case.

Assume then that $I(X^{n-1} \to Y^{n-1}) + I(0 * Y^{n-2} \to X^{n-1}) = I(X^{n-1}, Y^{n-1})$. Then, by definition,

$$\begin{aligned} I(X^n \to Y^n) &= \sum_{i=1}^{n} I(Y_i; X^i \mid Y^{i-1}) \\ &= I(X^{n-1} \to Y^{n-1}) + I(Y_n; X^n \mid Y^{n-1}) \end{aligned}$$

by simply separating the first $n-1$ terms of the sum from the last. Similarly,

$$\begin{aligned} I(0 * Y^{n-1} \to X^n) =& I(0 * Y^{n-2} \to X^{n-1}) \\ &+ I(X_n; Y^{n-1} \mid X^{n-1}). \end{aligned}$$

Applying the inductive hypothesis along with the identity $I(A, B; C) = I(A; C) + I(B; C \mid A)$ yields [11]

$$\begin{aligned} I(X^n \to Y^n) + I(0 * Y^{n-1} \to X^n) &= \\ &= I(X^{n-1}; Y^{n-1}) + I(Y_n; X^n \mid Y^{n-1}) + \\ &\quad + I(X_n; Y^{n-1} \mid X^{n-1}) \\ &= I(X^n; Y^{n-1}) + I(Y_n; X^n \mid Y^{n-1}) \\ &= I(X^n; Y^n). \end{aligned}$$

The concatenation of the 0 or null state before $Y^{n-1}$ is necessary as the canonical order in which the events happen is assumed to be $X_1, Y_1, X_2, \ldots, X_n, Y_n$; under this assumption $Y_n$ cannot influence $X_n$ because it has not happened yet.

Notice that there is nothing in the framework that prevents two processes from influencing each other; this conservation law shows that it is precisely when two-way influence is occurring that directed information is any more useful than mutual information.

For instance, it is known that $I(X^n; Y^n)$ is an upper bound of the information transfer over a communications channel. If the channel has feedback, however, some of the information transferred will be $I(0 * Y^{n-1} \to X^n)$ feedback. Hence a better upper bound of the capacity of the channel is $I(X^n \to Y^n)$ [9]. This is verified by the following equality, assuming $w$ is the source signal and $X_i = f(w, Y^{i-1})$:

$$\begin{aligned} I(w; Y^n) &= H(Y^n) - \sum_{i=1}^{n} H(Y_i \mid Y^{i-1}, w) \\ &= H(Y^n) - \sum_{i=1}^{n} H(Y_i \mid Y^{i-1}, X^i, w) \\ &= H(Y^n) - \sum_{i=1}^{n} H(Y_i \mid Y^{i-1}, X^i) \\ &= I(X^n \to Y^n). \end{aligned}$$

#### 2.2.2 Equivalence of Granger Causality and Directed Information in the Gaussian Case

Notice that if

$$\begin{aligned} X_{i+1} &= \alpha X_i + \xi_i \\ Y_{i+1} &= \beta Y_i + \gamma X_i + \zeta_i \\ &= \beta Y_i + \frac{\gamma}{\alpha} X_{i+1} - \frac{\xi_i}{\alpha} + \zeta_i \end{aligned}$$

then

$$\begin{aligned} I(X^n \to Y^n) &= \sum_{i=1}^{n} I(Y_i; X^i \mid Y^{i-1}) \\ &= \sum_{i=1}^{n} h(Y_i \mid Y^{i-1}) - h(Y^i \mid X^i, Y^{i-1}) \\ &= \sum_{i=1}^{n} \frac{1}{2} \log_2(2\pi e \epsilon^2) - \frac{1}{2} \log_2(2\pi e \tilde{\epsilon}^2) \\ &= n \log_2 \left( \frac{\epsilon}{\tilde{\epsilon}} \right). \end{aligned}$$

That is, directed information is a generalization of Granger causality that turns out to be equivalent in

the case where both processes are gaussian autoregressive. Recall that directed information has no assumptions about the distribution of $X$ and $Y$, while Granger causality implicitly assumes that $X$ and $Y$ are gaussian in the autoregression step.

## 2.3. Causal and direct causal influence

Directed information and Granger causality can both give false positives when used to determine if a direct causal link exists between two processes. These spurious influences fall into two broad categories. [10]

The first type of indirect influence is called a *proxy* influence [10], as in Figure 1. A process $X$ influences process $Z$ by proxy if $X$ influences process $Y$, process $Y$ in turn influences process $Z$, but all of $X$'s influence over $Z$ is explained by $Y \to Z$; that is, $X \mid Y$ is independent of $Z \mid Y$. This type of indirect influence is not entirely spurious as $X$ does have a real influence over $Z$; however it does imply a physical link that may not be present.

In fact, whether or not an influence is considered a proxy influence depends on what states are even relevant to the system. In a physical circuit, processes are often connected by metal wires that carry information in the form of electrical charge. A change in the state of some process, at the most basic level, could be said to have been "caused" by a single electron acting under the influence of some other process. This seems absurd, but the reason why is unclear—electrons ostensibly cannot make decisions, but the same could reasonably be said of any chip with deterministic dynamics. Where one draws the line depends on what meaning one wishes to extract from the nature of these events.

However, there are systems in which a multitude of state space "coarseness" are valid to some study. Saying, "Alice didn't shoot Bob, her gun did" would be a bad defense in court, but a gun legislature proponent might hypothesize that Bob would be alive were the gun not present. Moreover, the medical examiner who performs Bob's autopsy would very much be interested in exactly which bullet proved fatal, and the forensic analysts at the crime scene would look for the shell casings and gunpowder residue that "caused" this fatal bullet to obtain its lethal kinetic energy and trajectory. Yet at the end of the day, it is neither the bullet nor the gunpowder that gets sent to prison. Indeed, culpability is a very philosophical subset of causality that it is unlikely any formula or algorithm in the near future will be able to assign.

A more spurious form of indirect influence is *cascading* influence, as in Figure 2. This occurs when a process $X$ influences both processes $Y^{(1)}$ and $Y^{(2)}$, with some different time delay. Then, since both $Y^{(1)}$ and $Y^{(2)}$ have a mutual dependence on $X$, ignoring $X$ may lead to the appearance of influence between $Y^{(1)}$ and $Y^{(2)}$.

As an example, it might seem reasonable that lung cancer causes bad breath, but there may be no reason to believe that bad breath causes lung cancer. Smoking causes both lung cancer and bad breath; however since smoking causes bad breath in the short term, and lung cancer in the long term, one could reasonably imagine a dataset in which, if the presence of smoking is ignored, it appears that bad breath influences lung cancer.

In formalizing the concept of "influence", [10, eqns. 42-43] introduces two definitions. We say that the process $\{X_i\}$ **causally influences** $\{Y_i\}$ if

$$p_{Y^n||X^n} \neq p_Y, \tag{20}$$

that is,

$$\Pi p_{Y_i|Y^{i-1},X^i} \neq \Pi p_{Y_i|Y^{i-1}}. \tag{21}$$

The process $X$ causally influences the process $Y$ if and only if $I(X \to Y) > 0$. [10] The proof uses the KL-Divergence formulation of directed information, pointing out the fact that the KL-divergence is *positive definite*; that is, the KL-divergence between $p$ and $q$ is nonnegative, and zero if and only if $p = q$.

For the second definition, let $V = \{X^{(1)}, \ldots, X^{(m)}, Y\}$ be a set of processes. Then we say that $X^{(j)}$ **directly causally influences** $Y$ **with respect to** $V$ if for all $W \subset V \setminus \{X^{(j)}, Y\}$, $X^{(j)}$ causally influences $Y||W$.

This only determines if direct causal links exist; directed information should be used to quantify the nature of these causal links.

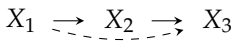


**Figure 1:** *$X_1$ influences $X_3$ by proxy*

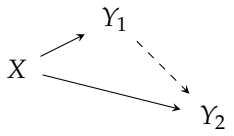


**Figure 2:** *Cascading influence from $Y_1$ to $Y_2$*

## 3. Applications

Directed information has a number of applications beyond Massey's motivation of studying feedback channels [9]. Among them are the following:

- Measuring the capacity of a channel with feedback [9]
- Mapping flow of information in the brain [10]
- Causal links between Twitter moods and stock market trends [2]
- Quantifying gambling/stock market strategy [1]
- Investigating if energy consumption causes economic growth [12]
- Characterizing protein interaction networks [13]

### 3.1. Directed Information in Gambling

Kelley et. al. [14] and Permuter et. al. [1] investigated the application of mutual- and directed-information respectively to quantify the value of "side information" when betting on a horse race.

Kelly showed in [14] that if each race outcome is an *i.i.d.* copy of some random variable $X$, and the gambler has side information $Y$ relevant to the outcome of $X$, then $I(X;Y)$ is the difference in growth rate of the gambler's portfolio; that is, if $S$ is the gambler's wealth before the race, and $SW$ is the gambler's wealth after the race without using side information, then $S(W + I(X;Y))$ is the gambler's wealth after the race if the side information is used optimally. [1]

Permuter [1] expanded on this idea by considering a sequence of horse races $X_1, X_2, \ldots X_n$ and some causal side information $Y_i$ that is relevant to the outcome of the race $X_i$. They show that the difference in growth rate between using and not using the side information $Y$ is $I(X^n \to Y^n)$, and that the normalized directed information reduces to $I(X;Y)$ when $X$ and $Y$ are *i.i.d.*, coinciding with Kelly's result.

The notations used are summarized:

- $X_i$ is the horse that wins race $i$;
- $Y_i$ is the side information known upon betting on race $i$;
- $o(x_i \mid x^{i-1})$ is the odds of $x_i$ winning race $i$ given all previous outcomes, and is the payoff to the gambler if $x_i = X_i$.
- $b(x)$ is the fraction of wealth invested in horse $x$. In particular, the paper frequently uses $b(X_i|Y^i, X^{i-1})$, the fraction invested on the *winning* horse given the knowledge of all side information and previous outcomes.
- $S(x^n||y^n)$ is the gambler's wealth if the race outcomes were $x^n$ and the information $y^n$ was causally available.
- $W(X^n||Y^n) = \mathbb{E}[\log(S(X^n||Y^n))]$ is the growth of wealth; $\frac{1}{n}W(X^n||Y^n)$ is the growth rate.

### 3.2. Twitter Moods and the Stock Market

The effects of public opinion have been hard to quantify, especially in real time. But with the advent of the social media site Twitter, researchers have been granted access to social data that allows them to demonstrate precisely how social exchanges influence the world at large.

An example is the prediction of stock market prices based on Twitter mood. It has long been assumed that the stock market behaves in a stochastic manner, influenced only by current events, but work by Bollen, Mao, and Zeng have demonstrated that certain mood trends can be used to accurately predict closing prices

of the Dow Jones Industrial Average (DJIA) [2]. The procedure was to aggregate millions of tweets sent between February 28th to December 19th 2008, and analyze their content using two tools: OpinionFinder and GPOMS. OpinionFinder was used to extract how "positive" or "negative" a series of tweets were and GPOMS was used to separate their emotions into six district categories: Calm, Alert, Sure, Vital, Kind, and Happy.

To determine whether or not Twitter mood contained predictive information about the DJIA, two linear regressions were performed in accordance with granger causality analysis.

The first of which was a simple time-lagged regression for day-to-day DJIA changes,

$$D_t = \sum_{i=1}^{n} \epsilon_i D_{t-i} + \epsilon_t \tag{22}$$

Where $D_t$ is the the change in DJIA between days $t$ and $t-1$, and $n$ is the number of previous days taken into consideration.

The second analysis included the Twitter mood data obtained by GPOMS:

$$\widetilde{D}_t = \sum_{i=1}^{n} \epsilon_i D_{t-i} + \widetilde{\epsilon}_t \tag{23}$$

It was found that the Calm time series demonstrated the most granger causal relation to DJIA changes [2]. Intuitively, this makes sense, as calm traders probably make for more stable market prices. This was shown by huge shift in the DJIA and the Calm series during the bank bails outs.

Figure 3

According to this paper, these results present an addition to an assumption of the Efficient Market Hypothesis, which implies that market prices will be driven by news, and therefore should be random [2]. As a conclusion, it appears that the stock market changes are also driven by mood, particularly by general calmness.

## 4. Conclusions

Despite its philosophical controversy, perhaps nothing more than a result of its controversial name, Granger causality has opened the door to a new wave of statistical analysis beyond that offered by simple regression. We've shown the insight it provides into our understanding of the links between prevailing moods in society and their effect on the stock market. Granger causality provides similar insight into the flows of information between other processes as well. Directed information generalizes this idea to processes that are not gaussian autoregressive, and has physical meaning as the increase in growth rate of a portfolio when side information is known, as well as the capacity of a communication channel with feedback. With conditioning, these tools can also determine the flow of information between two processes that isn't due to some other process. However, the study of causality does require some thought into the definition of the state space, in regards to what kind of meaning one wishes to extract.

## References

[1] H. H. Permuter, Y.-H. Kim, and T. Weissman, "On directed information and gambling," in *Information Theory, 2008. ISIT 2008. IEEE International Symposium on*. IEEE, 2008, pp. 1403–1407.

[2] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *CoRR*, vol. abs/1010.3003, 2010. [Online]. Available: http://arxiv.org/abs/1010.3003

[3] G. Kramer, "Directed information for channels with feedback," Ph.D. dissertation, University of Manitoba, Canada, 1998.

[4] P. Suppes, *A probabilistic theory of causality*, ser. Acta philosophica Fennica. North-Holland Pub. Co., 1970. [Online]. Available: http://books.google.com/books?id=Ff4HAQAAIAAJ

[5] C. Granger, "Some recent development in a concept of causality," *Journal of Econometrics*, vol. 39, no. 1âĂŞ2, pp. 199 – 211, 1988. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0304407688900450

[6] C. W. J. Granger, "Testing for causality : A personal viewpoint," *Journal of Economic Dynamics and Control*, vol. 2, no. 1, pp. 329–352, May

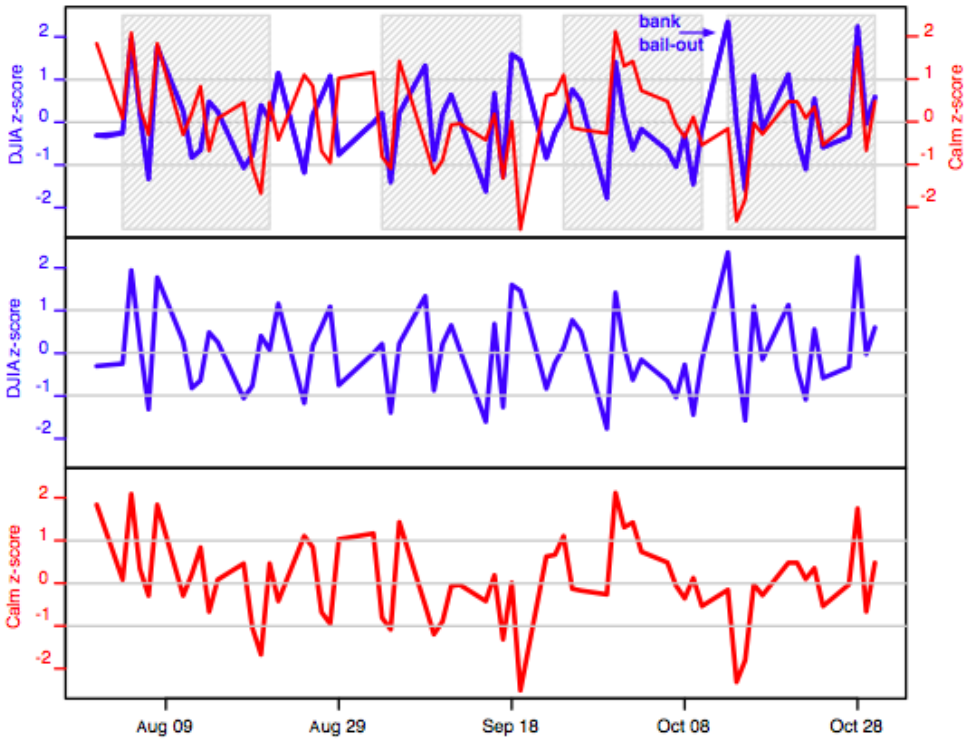


**Figure 3:** *"Z-scores comparing the DJIA changes and the Calm time series. Bank bail-out peak is shown on the right. Image from Bollen, Mao, and Zeng. "*

1980. [Online]. Available: http://ideas.repec.org/a/eee/dyncon/v2y1980i1p329-352.html

[7] M. Ding, Y. Chen, and S. L. Bressler, "17 granger causality: Basic theory and application to neuroscience," *Handbook of time series analysis: recent theoretical developments and applications*, p. 437, 2006.

[8] H. Marko, "The bidirectional communication theory–a generalization of information theory," *Communications, IEEE Transactions on*, vol. 21, no. 12, pp. 1345–1351, 1973.

[9] J. Massey, "Causality, feedback and directed information," in *Proc. Int. Symp. Inf. Theory Applic.(ISITA-90)*. Citeseer, 1990, pp. 303–305.

[10] C. J. Quinn, T. P. Coleman, N. Kiyavash, and N. G. Hatsopoulos, "Estimating the directed information to infer causal relationships in ensemble neural spike train recordings," *Journal of computational neuroscience*, vol. 30, no. 1, pp. 17–44, 2011.

[11] J. L. Massey and P. C. Massey, "Conservation of mutual and directed information," in *Information Theory, 2005. ISIT 2005. Proceedings. International Symposium on*. IEEE, 2005, pp. 157–158.

[12] G. Hondroyiannis, S. Lolos, and E. Papapetrou, "Energy consumption and economic growth: assessing the evidence from greece," *Energy Economics*, vol. 24, no. 4, pp. 319–336, 2002.

[13] K. Sachs, O. Perez, D. Pe'er, D. A. Lauffenburger, and G. P. Nolan, "Causal protein-signaling networks derived from multiparameter single-cell data," *Science*, vol. 308, no. 5721, pp. 523–529, 2005.

[14] J. L. Kelly, "A new interpretation of information rate," *Information Theory, IRE Transactions on*, vol. 2, no. 3, pp. 185–189, 1956.

[15] C. W. Granger, "Investigating causal relations by econometric models and cross-spectral methods,"

*Econometrica: Journal of the Econometric Society*, pp. 424–438, 1969.

[16] J. Rissanen and M. Wax, "Measures of mutual and causal dependence between two time series (corresp.)," *Information Theory, IEEE Transactions on*, vol. 33, no. 4, pp. 598–601, 1987.

[17] X. Wen, G. Rangarajan, and M. Ding, "Is granger causality a viable technique for analyzing fmri data?" *PLoS ONE*, vol. 8, no. 7, p. e67428, 07 2013. [Online]. Available: http://dx.doi.org/10.1371%2Fjournal.pone.0067428

[18] P. Foresti, "Testing for Granger causality between stock prices and economic growth," University Library of Munich, Germany, MPRA Paper 2962, 2006. [Online]. Available: http://ideas.repec.org/p/pra/mprapa/2962.html

[19] E. T. Rosenman, "Retweets—but not just retweets: Quantifying and predicting influence on twitter," Ph.D. dissertation, 2012.