This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3068273, IEEE Access

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

Panel data causal inference using a rigorous information flow analysis for homogeneous, independent and identically distributed datasets

Yineng Rong[12], **X. San Liang**[12*]

[1]School of Atmospheric Sciences, Nanjing University of Information Science and Technology, Nanjing 210044 China
[2]Center for Ocean-Atmosphere Dynamical Studies, Nanjing University of Information Science and Technology, Nanjing 210044 China

*Corresponding author: X. San Liang (e-mail: x.san.liang@gmail.com).

This work was supported in part by National Science Foundation of China under Grant 41975064 and the 2015 Jiangsu Program for Innovation Research and Entrepreneurship Groups.

ABSTRACT Panel data, which consist of observations on many individual units over two or more instances of time, have gradually become an important type of scientific data. Subsequently causal inference for panel data has attracted enormous interest from many fields as well as statistics. In this study, the rigorously formulated information flow analysis for time series, which is very concise in form and has been successfully applied in different disciplines, is generalized to identify the causality from homogeneous and independent identically distributed panel data. The resulting formula bears the same form as that for the former, though the meanings of the symbols differ. An algorithm is then proposed for panel data causality analysis, which has been validated with both linear and nonlinear problems. It has also been put to application to examine the causal relations among economic growth, energy consumption, trade openness, and energy price based on 15 Asian countries. Clearly identified are a strong bidirectional causality between economic growth and energy consumption, and a strong causality from import and export trade to economic growth. Energy price has no direct impact on energy consumption; it, instead, exerts a weak effect on the latter through influencing economic growth.

INDEX TERMS Panel data, Information flow, Causality, Economy, Energy

I. INTRODUCTION

In the past two decades, data have been accumulated at an exponential rate in essentially all fields, partly due to the easy access to social media and the interconnectivity of our society [1]. How to mine the causal information from the different datasets hence becomes a hot issue in the digitized society [2]. One direct way is to identify the causal possible relations. Unfortunately, causal inference is a very challenging problem. So far as of today, the methodologies for identifying causality are yet to be improved[3].

Most of the data can be classified into three categories: temporal data, cross-sectional data, and panel data. A set of temporal data or time series is a series of data points indexed (or listed or graphed) in time order. Differently, data collected by observing many individuals at the one instance of time is termed cross-sectional. Time series and cross-sectional data can be thought of as special cases of panel data, which consist of observations on many individual units over two or more periods of time. There are several important advantages of panel data comparing to data sets with only a temporal (time series) or individual (cross section) dimension [4], one being the ability to control for possibly correlated, time-invariant heterogeneity without actually observing it. Besides, panel data can reduce the collinearity among explanatory variables, increase in efficiency of estimators, and alleviate problems of aggregation.

Several methods have been proposed to make causal inference with panel data, among which the most popular one is Granger causality analysis, which is based on the idea that the cause occurs before the effect, hence if an event X is the cause of another event Y, then X should proceed Y [5]. (This basis, however, is recently challenged by an observation with a purported generated dynamical system with synchronization; see [6]). For example, Holtz-Eakin et. al. [7] considered estimation and testing of vector autoregression (VAR) coefficients in panel data to calculate the Granger causality, and applied the techniques to analyze the dynamic

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

relationships between wages and hours worked in two samples of American males. The same method was used by [8] to report the findings on the relationship between foreign direct investment and pollution across 112 countries over 15–28 years. Kónya [9] used the Panel-Data Granger causality test approach based on bootstrap and studied the relationship between exports and economic growth in OECD countries. Similar approach was adopted by Bedir and Yilmaz[10] to examine the causal relation between the logarithms of the human development index and CO_2 emissions in 33 organizations for economic cooperation and development countries. Gupta and Singh [11] employed the Johansen cointegration technique followed by vector error correction model (VECM) and standard Granger causality test to investigate the causal linkage between FDI and GDP of BRICS nations. These applications are generally successful in their respective contexts.

Despite these studies, the causality analysis for panel data is still in its early stage of development. Both theoretically and practically there still exist much room for improvement. Recently, it has been realized that causality and information flow (IF) are real physical notions and hence can be put on a rigorous footing. In other words, they can be derived from first principles in physics [12], rather than axiomatically proposed as an ansatz.

Effort along this line can be traced back to the early work by Liang and Kleeman [13] on IF, but its ability has just been recognized with the publication of the time series study by Liang [14], where it is shown that causality can be assessed in a very easy way, with only sample covariances involved. The resulting formula, albeit simple, proves to be remarkably successful in solving many problems which defy the traditional approaches. It also fixes the philosophical debate on causation versus correlation (cf. section 2). Ever since then, the IF-based causality analysis has been widely applied to the problems with time series such as global warming [15], [16], El Niño [14], typhoon genesis prediction [17], space weather [18], chlorophyll variability [19], relation between soil moisture and precipitation [20], financial time series analysis [21], neuroscience problems [22], to name a few.

Considering the success of the IF-based causality analysis for time series, we henceforth want to generalize it to panel data. In the following a brief introduction of the theory is first presented, then in section III, we show that a generalization can be fulfilled, and an algorithm is then proposed. In section IV, the algorithm is validated with a linear stochastic system and a highly chaotic deterministic system. Section V give an application and section VI summarizes the whole study.

II. INFORMATION-FLOW AND CAUSALITY BETWEEN TIME SERIES—A BRIEF REVIEW

Different from the various statistical approaches for causal inference, the information flow-based causality analysis is derived from first principles in physics. Ever since Liang and Kleeman [13], much effort has been invested to establish a rigorous formalism which has just been fulfilled [12]. Accordingly a causal inference technique is developed for time series [14]. It is concise in form, easy to implement and, moreover, quantitative in nature (see below (3)). Since its advent, many applications in different disciplines have been carried out with remarkable success. The following material is just a very brief introduction of the theory that is needed for this study. For a systematic treatment and other materials, see [12], among other papers.

This line of work begins with the concept of information flow which is defined as follows:

Definition II.1 In a dynamical system (Ω, Φ_t) where Ω is the phase space and Φ_t may be a flow or a discrete mapping, the information flow from a component X_2 to another component X_1, written $T_{2\to1}$, is defined as the contribution of entropy from X_2 per unit time (continuous time case) or per step (discrete mapping case) in increasing the marginal entropy of X_1 as the state is steered forth by Φ_t.

With this, causality can be defined, in a quantitative sense,

Definition II.2 X_2 is causal to X_1 iff the information flow $T_{2\to1} \neq 0$. The strength of the causality from X_2 to X_1 is measured by $|T_{2\to1}|$. Likewise, the causality from X_1 to X_2 can be defined.

Remark 1. A nonzero $T_{2\to1}$ may be either positive or negative. A positive $T_{2\to1}$ means that X_2 makes X_1 more uncertain, and vice versa. But for the purpose of causal inference, the sign is not essential; we just consider its magnitude.

Remark 2. By the definition, we can distinguish three cases: (1) noncausal ($T_{2\to1} = T_{1\to2} = 0$), (2) unidirectionally causal ($T_{2\to1} \neq 0$, $T_{1\to2} = 0$ or $T_{2\to1} = 0$, $T_{1\to2} \neq 0$), (3) bidirectionally causal ($T_{2\to1} \neq 0, T_{1\to2} \neq 0$), as discussed in [23].

Remark 3. In the above definitions entropy is generally understood as Shannon entropy, but other entropies may also apply. In this study, we stick to Shannon entropy.

Now consider a two-dimensional (2D) stochastic dynamical system

$$d\boldsymbol{X} = \boldsymbol{F}(\boldsymbol{X}, t)dt + \boldsymbol{B}(\boldsymbol{X}, t)d\boldsymbol{W}, \tag{1}$$

where $\boldsymbol{F} = (F_1, F_2)$ is the vector of drift coefficients, $\boldsymbol{X} = (X_1, X_2) \in \mathbb{R}^2$ are the random variables, $\boldsymbol{W} = (W_1, W_2)$ is a standard 2D Wiener process and $\boldsymbol{B} = (b_{ij})$ is the matrix of diffusion/volatility coefficients. Liang [24] established that the time rate of IF from X_2 to X_1 with respect to Shannon entropy is:

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3068273, IEEE Access

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

where ρ is the joint probability density function of $\boldsymbol{X}$, ρ_1 is the marginal density of X_1, $g_{11} = \sum_{k=1}^{2} b_{1k}^2$, and E is the expectation with respect to ρ. Later on it has been shown that the formula is the same with respect to Kullback-Leiber divergence [25]. Likewise, the IF from X_1 to X_2 is

$$T_{1\to2} = -E\left(\frac{1}{\rho_2}\frac{\partial F_2\rho_2}{\partial x_2}\right) + \frac{1}{2}E\left(\frac{1}{\rho_2}\frac{\partial^2 g_{22}\rho_2}{\partial x_2^2}\right),$$

Ideally, if $T_{2\to1} = 0$, then X_2 is not causal to X_1; otherwise it is causal, and the magnitude of $|T_{2\to1}|$ means the strength of the causality. The larger $|T_{2\to1}|$, the stronger causality from X_2 to X_1.In practice, significance should be tested prior to making the inference.

The above derived information flow has many important properties. The first is the "Principle of Nil Causality" [12]: a process, say X, has a zero causality to another process, say Y, if the evolution of Y does not depend on X. This is a basic principle that all formalisms try to verify in applications, while in this formalism, it is a proven theorem. Many other properties can be seen in [12] and [25].

The IF formula has been validated with many highly chaotic systems, such as baker transformation, Hénon map, Kaplan-Yorke map, Rössler system, truncated Burgers-Hopft system, to name a few [12], [26]. Under a linearity assumption, Liang[14] further established that it can be estimated from two time series, say, X_1 and X_2. The resulting maximum likelihood estimator is:

$$\hat{T}_{\ 2\to1} = \frac{C_{11}C_{12}C_{2,d1} - C_{12}^2 C_{1,d1}}{C_{11}^2 C_{22} - C_{11}C_{12}^2}, \qquad (3)$$

where C_{ij} is the covariance between X_i and X_j, $C_{i,dj}$ is the covariance between X_i and $\dot{X}_j$, and $\dot{X}_j = (X_j(t+k\Delta t) - X_j(t))/(k\Delta t)$ is the difference approximation of dX_j/dt using the Euler forward scheme. Here k is usually 1; for cases of deterministic chaos, it should be set 2. This formula is very simple in form but evidently very successful in real applications, some of which have been mentioned in the introduction above.

Considering that there is a long-standing philosophical debate over causation versus correlation, rewrite (3) in terms of correlation coefficients:

$$\hat{T}_{2\to1} = \frac{r}{1-r^2}\left(r'_{2,d1} - r r'_{1,d1}\right), \qquad (4)$$

here $r = \frac{C_{11}}{\sqrt{C_{11}C_{22}}}$ is the sample correlation coefficient between X_1 and X_2, and $r'_{i,dj} = \frac{C_{i,dj}}{\sqrt{C_{11}C_{22}}}$ the "correlation coefficient" between X_{i} and $\dot{X}_j$. So, if $r = 0, \hat{T}_{2\to1} = 0$; the converse, however, is not necessarily true. In other words, **causation implies correlation, but correlation does not imply causation.** Equation (4), therefore, bridges causation and correlation with a simple mathematical relation.

III. CAUSALITY ANALYSIS FOR HOMOGENEOUS I.I.D. PANEL DATA -– AN ALGORITHM

Panel data not only consist of observations over time, but also over many individual units. The above dynamical system-based formula then may not be directly applicable. This is different from Granger causality, which is fundamentally a notion of probabilistic conditional independence, and hence can be applied not only to time series data but also to cross-section and panel data [27]. We need to re-establish from scratch a formula of the like of (3). We first give a definition for panel data causality.

Definition III.1 For a homogeneous panel dataset, the causality from a variable, say X_2, to another variable X_1 between two cross-sections is defined as the absolute value of the information flow from X_2 to X_1 as the underlying system evolves between the two steps.

Remark: For panel datasets with more than 2 cross-sections, a relation of causality vs. time step can be obtained by computing the information flows between adjacent steps.

As Liang [14], we assume a linear model. Though this sets a limitation, the formula (3) has proved to be remarkably successful in many highly nonlinear problems. In fact, this is not surprising; anyhow, when correlation is referred we usually mean linear correlation.

Theorem III.1 Suppose a homogeneous i.i.d. panel dataset is generated through some linear system with Wiener processes, and X_2 and X_1 are the two variables of the dataset. Then the information flow from X_2 to X_1 between two adjacent steps $(t, t+\Delta t)$ is

$$T_{2\to1} = \frac{\sigma_{12}}{\sigma_{11}} \cdot \frac{(-\sigma_{12}\sigma_{1,d1} + \sigma_{11}\sigma_{2,d1})}{(\sigma_{11}\sigma_{22} - \sigma_{12}^2)}, \qquad (5)$$

where σ_{ij} are population covariances between X_i (t) and $X_j(t)$, and $\sigma_{i,dj}$ $=E(X_i(t) - EX_i(t))(\Delta X_j - E\Delta X_j)/\Delta t$, with $\Delta X_j = X_j(t+\Delta t) - X_j(t)$.

Proof

In (1), let

$$\boldsymbol{F} = \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} = \begin{bmatrix} a_{11}X_1 + a_{21}X_2 \\ a_{12}X_1 + a_{22}X_2 \end{bmatrix}. \qquad (6)$$

It has been established in [14] that (2) is reduced to

$$T_{2\to1} = a_{11}\frac{\sigma_{12}}{\sigma_{11}}, \qquad (7)$$

which is remarkably simple. Here σ_{ij} make the entries of the population covariance matrix. We now estimate this formula, given an individual independent ensemble of panel data with two time instances spanned by an interval Δt.

Different from the time series considered in [14], which requires some extra assumption such as stationary, here the estimation of (7) turns out to be much easier. The reason is that (2) appears in a form of ensemble mean, while a set of panel data provides a natural ensemble. As Liang[14], discretize (1) with the Euler-Bernstein scheme the dynamical system to get

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3068273, IEEE Access

$$X_1(t+\Delta t) = X_1(t) + \Delta t(a_{11}X_1 + a_{21}X_2) + b_{11}\Delta W,$$

where $\Delta w \sim \mathcal{N}(0, \Delta t)$. For convenience, rewrite it as

$$a_{11}X_1 + a_{21}X_2 = \Delta X_1/\Delta t - b_{11}\Delta W/\Delta t. \qquad (8)$$

Considering the availability of the ensemble, take ensemble mean to get

$$Ea_{11}X_1 + Ea_{21}X_2 = E(\Delta X_1/\Delta t). \qquad (9)$$

Subtracting (9) from (8), then multiplying by $(X_1 - EX_1)$, and taking expectation, we get

$$E(X_1 - EX_1)(X_1 - EX_1)a_{11} + E(X_1 - EX_1)(X_2 - EX_1)a_{12}$$
$$= E(X_1 - EX_1)(\Delta X_1 - E\Delta X_1)/\Delta t + 0.$$

This is

$$\sigma_{11}a_{11} + \sigma_{12}a_{12} = \sigma_{1,d1}, \qquad (10)$$

where σ_{ij} are covariances between X_i and X_j, and $\sigma_{i,dj}$ $= E(X_i - EX_i)\left(\Delta X_j - E\Delta X_j\right)/\Delta t$. Likewise, the difference between (8) and (9) multiplied by $(X_2 - EX_2)$, followed by a mathematical expectation results in

$$\sigma_{12}a_{11} + \sigma_{22}a_{12} = \sigma_{2,d1}, \qquad (11)$$

(10) and (11) combined to give

$$a_{12} = \frac{-\sigma_{12}\sigma_{1,d1} + \sigma_{11}\sigma_{2,d1}}{\sigma_{11}\sigma_{22} - \sigma_{12}^2}. \qquad (12)$$

Substitute back to (7) and we get

$$T_{2\to1} = \frac{\sigma_{12}}{\sigma_{11}} \cdot \frac{(-\sigma_{12}\sigma_{1,d1} + \sigma_{11}\sigma_{2,d1})}{(\sigma_{11}\sigma_{22} - \sigma_{12}^2)}.$$

Q.E.D.

In real applications, the population covariances need to be replaced with sample covariances. This results in a formula

$$T_{2\to1} = \frac{C_{12}(C_{11}C_{2,d1} - C_{12}C_{1,d1})}{C_{11}(C_{11}C_{22} - C_{12}C_{12})}, \qquad (13)$$

which is in a form precisely the same as (3), except that now the mean is ensemble mean at time t, not time average. Here $T_{2\to1}$ is understood as an estimator, and should have been written $\hat{T}_{2\to1}$, but for simplicity, the hat is dropped. Similarly, the IF from X_1 to X_2 is

$$T_{1\to2} = \frac{C_{12}(C_{22}C_{1,d2} - C_{12}C_{2,d2})}{C_{22}(C_{11}C_{22} - C_{12}C_{12})}, \qquad (14)$$

which is absolutely different from (13). This naturally indicates the direction of causality. If the absolute value of $T_{2\to1}$ ($|T_{2\to1}|$) passes the significance test, it is believed that X_2 is the cause of X_1. Similarly, if $|T_{1\to2}|$ passes the test, X_1 is the cause of X_2.

When there are multiple time steps, say K steps, (13) may be applied to each two adjacent time instances, and hence obtain $(K-1)$ information flows, over which an average information flow result. We hence have the following algorithm.

Algorithm-IF: Information flow for homogeneous i.i.d. panel data

```
Input: Panel data $X_{p1}$ and $X_{p2}$ with dimension $N \times K$, where $N$ is the number of individual units and $K$ the time steps, every two adjacent steps separated by a time interval $\Delta t$.
Step 1: Let $X_1$, $X_2$ and $\dot{X}_1$ be three empty 1D vectors.
for $i = 1{:}N$
    for $j = 2{:}K$
        $\mathrm{tmp} = (X_{p1}(i,j) - X_{p1}(i,j-1))/\Delta t)$;
        $\dot{X}_1 = [\dot{X}_1; \mathrm{tmp}]$;
        $X_1 = [X_1\,;\, X_{p1}(i,j-1)]$;
        $X_2 = [X_2\,;\, X_{p2}(i,j-1)]$;
    end
end
note: $[A; B]$: concatenates $B$ vertically to the end of $A$.
Step 2: Calculate the covariances: $C_{11}$, $C_{12}$, $C_{22}$, $C_{1,d1}$ and $C_{2,d1}$ with $X_1$, $X_2$, and $\dot{X}_1$.
Step 3: Substitute the covariances into (13) and (14) to get the information flow from $X_2$ to $X_1$ ($T_{2\to1}$) and that from $X_1$ to $X_2$ ($T_{1\to2}$).
Output: $T_{2\to1}$, $T_{1\to2}$.
```

IV. Validation

A. A linear problem

We first use a discretized version of (1) to generate a set of panel data. Assuming that $\boldsymbol{F}$ and $\boldsymbol{B}$ have the following form

$$\boldsymbol{F} = \begin{bmatrix} a_{11}X_1 + a_{21}X_2 \\ a_{12}X_1 + a_{22}X_2 \end{bmatrix} = \begin{bmatrix} 0.3X_1 + 0X_2 \\ 0.5X_1 + 0.7X_2 \end{bmatrix},$$

$$\boldsymbol{B} = \begin{bmatrix} 0.4 & 0 \\ 0 & 0.5 \end{bmatrix}.$$

Choose $\Delta t = 0.01$, and hence $\Delta W = \sqrt{\Delta t}\, R_{\mathcal{N}}$, where $R_{\mathcal{N}}$ is a random number satisfied the standard normally distribution. This forms a 2D autoregressive process. Clearly, X_1 causes X_2, but not vice versa. This kind of problem is usually used to verify a causality analysis: One component causes another, but the latter does not cause the former. We initialize the system by making 10000 draws as follows:

$$X(t=0) = \begin{cases} X_{1,t=0} = 0.3 + 0.1R_{\mathcal{N}} \\ X_{2,t=0} = 0.4 + 0.1R_{\mathcal{N}} \end{cases}.$$

Fig. 1a shows that the initial distribution of X_1 and X_2 roughly meet the normal distribution of:

$$\mathcal{N}\left(\begin{pmatrix} 0.3 \\ 0.4 \end{pmatrix}, \begin{bmatrix} 0.1^2 & 0 \\ 0 & 0.1^2 \end{bmatrix}\right).$$

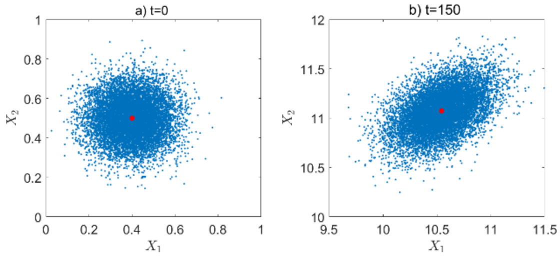

Fig. 1. a) The initial distribution (blue spots) and the ensemble mean (red spot) of X_1 and X_2. b) The distribution of X_1 and X_2 at $t = 150$ unit time.

For each initial condition the system is integrated for 15,000 steps, and the resulting X_1 and X_2 are recorded, and eventually form the ensemble. When $t = 150$ (Fig. 1b), the distribution has been inclined along the direction of $X_2 = X_1$, which means that X_1 and X_2 are no longer independent. Fig. 2 is a typical series with the initial condition of $X_{1,t=0} = 0.3$ and $X_{2,t=0} = 0.4$. After t=10, the system reaches a quasi-stationary state. We hence discard the segment $t < 10$ in forming the panel data.

According to the size of ensemble or number of individual units (N), and temporal series (K), panel data are

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3068273, IEEE Access

generally divided into three categories: the 'large K, small N' temporal style long sequences; the 'small K, large N' panel literature, and the 'large K, large N' heterogenous panel data [4]. By the assumptions in Theorem III.1, here the heterogeneous case is excluded. Based on this we henceforth generate three datasets, and respectively calculate the causalities between X_1 and X_2.

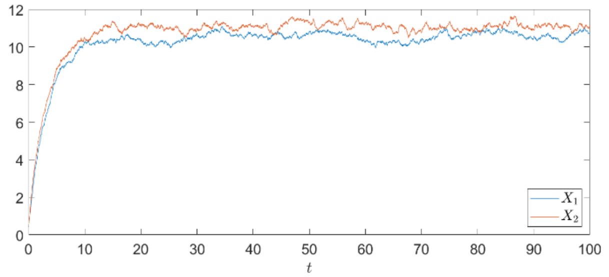

Fig. 2. A typical series generated by the 2D autoregressive process initialized with $X_1 = 0.3$ and $X_2 = 0.4$.

Case 1. The single pair time series case as studied in Liang (2014). Considers one realization over the period of $t = 50 - 150$ (steps from 5000 to 15000). The information flow is computed using (3). (Now **Algorithm-IF** boils down to this time series analysis case.)

Case 2. Generate a total of 10000 pairs of X at the section $t = (100 - \Delta t)$ and $t = 100$. Compute the information flow using **Algorithm-IF**.

Case 3. Generate 100 pairs of X over the period $t = (100 - 101\Delta t) - 100$, for 100 steps with equal time stepsize Δt. Compute the information flow using **Algorithm-IF**.

TABLE 1
INFORMATION FLOWS COMPUTED WITH THE 7 SETS OF PANEL DATA AS GENERATED.

	Case 1	Case 2	Case 3
$\lvert T_{2\to1}\rvert$	0.0176	0.0169	0.0110
$\lvert T_{1\to2}\rvert$	0.3501	0.2100	0.2996

It is found that in all these cases, $|T_{1\to2}|$ is nearly an order of magnitude larger than $|T_{2\to1}|$. Further, we adopted the same significance test as [14]. $|T_{1\to2}|$ in the 3 cases are all passes the 99% significance test, while there are no cases that $|T_{2\to1}|$ passes even the 80% significance test. We remark that in Case 1, using **Algorithm-IF** or (3) gives exactly the same result, indicating that time series data is just a particular case of panel data. By Table 1, **Algorithm-IF** for these panel data is robust.

B. A highly nonlinear problem

In deriving (13), a linear assumption is invoked. That is to say, strictly speaking, **Algorithm-IF** computes linear causality. Since (3) has been evidenced remarkably successful in highly nonlinear problems, we here test (13) and **Algorithm-IF** with such a dataset.

The panel data set is generated with a one-way coupled *anticipatory* map. This is a highly chaotic system designed by Hahs and Pethel [28] which fails the existing causal inference techniques then:

$$\begin{aligned} &X_1(t+1) = f\big(X_1(t)\big), \\ &X_2(t+1) = (1-\varepsilon) f\big(X_2(t)\big) + \varepsilon g_\alpha\big(X_1(t)\big), \end{aligned} \qquad (14)$$

where,

$$\begin{aligned} &f(x) = 4x(1-x), \\ &g_\alpha(x) = (1-\alpha) f(x) + \alpha f^2(x), \end{aligned}$$

and f^2 means that the logistic map f applies twice, α is a parameter called the "anticipation parameter". Picking $\varepsilon = 0.3$, α = 0.8, an example series pair is shown in Fig. 3. From (14) obviously X_1 causes X_2, but not vice versa. However, Hahs and Pethel [28] showed that, with the existing technique, the causality thus inferred becomes widely off the track as α increases on $\alpha \in [0,1]$. When $\alpha > 0.5$, not only the computed causality from X_2 to X_1 becomes dominating that from the other way around. We hence generate some panel data sets with this touch-stone system to test our algorithm. The anticipation parameter α takes value from 0 to 1 every 0.1. Like the linear runs for each α, with the initial conditions as:

$$\boldsymbol{X}(t=0) = \begin{cases} X_1 = 0.4(1 + 0.1R_N) \\ X_2 = 0.1(1 + 0.1R_N) \end{cases}.$$

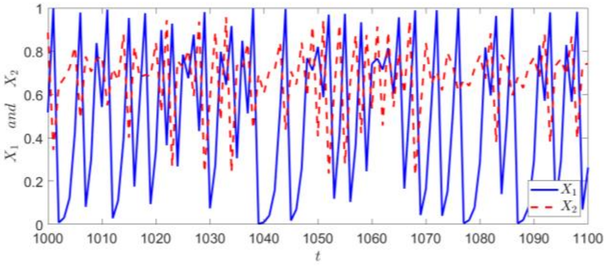

Fig. 3. Time series for $\alpha = 0.8$ with the initial condition $X_1 = 0.4$ and $X_2 = 0.1$.

For each group of runs, the system is iterated by 10,000 times, when the resulting X_1 and X_2 are recorded.We check three cases with this map: cases 4, 5, 6, which are the same as cases 1, 2, 3, respectively, but with the nonlinear anticipatory system. The two time steps for case 5 are 9998 and 10000, respectively. Fig. 4 is the absolute value of the information flow ($|T_{2\to1}|$ and $|T_{1\to2}|$) under the different cases and different anticipation parameters. The information flows with the panel data (no matter with large N, small K or the large N, large K) are similar as the result of the time series information flow as obtained by Liang (2014). Most importantly, $|T_{2\to1}|$ is very small throughout, though not exactly zero (perhaps due to the linear model used). Secondly, for $0 \le \alpha \le 0.3$ or $0.8 \le \alpha \le 1.0$, $|T_{1\to2}|$ is much larger than $|T_{2\to1}|$, indicating a one-way causality in a consistent way. This is in sharp contrast to the counterintuitive result of spurious causality as discovered by Hahs and Pethel [28].

When $0.4 \le \alpha \le 0.7$, the information flow from X_1 to X_2 is quite small. But even in such situations, the $|T_{1\to2}|/|T_{2\to1}|$ in all the cases are all no less than 1.5, and, besides, $T_{1\to2}$ passes the 99% significance test, while $T_{2\to1}$ does not pass the 95% significance test. In a word, though with a linear assumption, **Algorithm-IF** can capture the

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

causality among an otherwise highly nonlinear panel dataset in a consistent way.

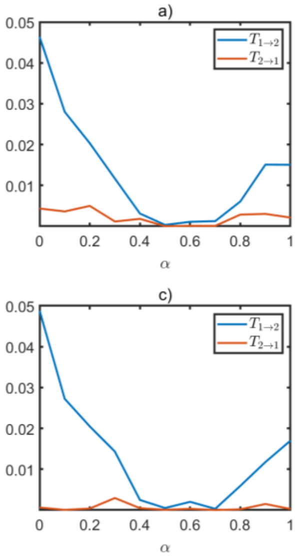

Fig. 4. The absolute value of IF (units: nats/unit time) from X_1 to X_2, $|\mathrm{T}_{1\to2}|$ (blue line) and that from X_2 to X_1, $|\mathrm{T}_{2\to1}|$ (red line) under different anticipation parameter α. a) Case 4; b) Case 5; c) Case 6. The result here is in sharp contrast to the classical ones, in which the red line dominates, as shown in Hahs and Pethel[28].

V. Real problem application

So far panel data are mostly investigated in economics. For this reason, we apply **Algorithm-IF** to a problem on economy versus energy. Specifically, it is about the causal relationship between economic growth, trade openness and energy consumption, based on the data of 15 Asian countries (Pakistan, India, Bangladesh, Sri Lanka, Philippines, Thailand, Indonesia, China, Malaysia, Japan, Jordan, Iran, South Korea, Nepal and Vietnam) over the period of 1980–2011. The problem has been studied by [29], hereafter NA14. They found a bi-directional causality among the above four factors (Table 7 in NA14). Here, we re-examine the problem with the above proposed new algorithm based on a rigorously developed theory.

A detailed description of the data is referred to NA14. Briefly, energy consumption is measured by the Kg of oil equivalent per capita; economic growth is by real GDP per capita in constant international dollar; exports (US$) plus imports (US$) divided by population is used to measure trade openness; the price of Dubai crude oil (US$) deflated by the country's consumer price index (100 in the year of 2005) is used as a proxy for energy price due to the unavailability of energy price data. Data on energy consumption per capita, merchandise exports, merchandise imports, consumer price index and population are obtained from World Development Indicators (2013) of the World Bank. Data on real GDP per capita are collected from Penn World Tables Version 8.0 [30] and Dubai crude oil price data are taken from British Petroleum's 2013 statistical review of world energy[31].

We calculate the information flows/causalities among the four factors with our **Algorithm-IF**. Similar to the Granger causalities as computed in NA14, we regard the causality with a p-value of information flow less than 0.05, 0.10, 0.15 as, respectively, strong, normal, and weak causality. The results are tabulated in Table 2, with information flows significant at an 85% confidence level blackened. For easy illustration, the causal relation is summarized in Fig. 5. From its economic growth and energy consumption are mutually causal, but the causality between economic growth and trade openness, and that between economic growth and energy price are one-way. Specifically, there is a strong bidirectional causality between economic growth and energy consumption, a strong unidirectional causality from trade openness to economic growth, and a weak unidirectional causality from energy price to economic growth. The first two are significant at a 99% confidence level; the third is significant at an 85% level. All other causalities (in total there could be $4 \times 3 = 12$ causalities) have not passed the significance test at the 85% confidence level, particularly, energy price (oil price) has no direct causal relationship with either energy consumption or trade openness, though it does exert a limited impact on the economic growth (significant at 85% confidence level; indicated by dashed line).

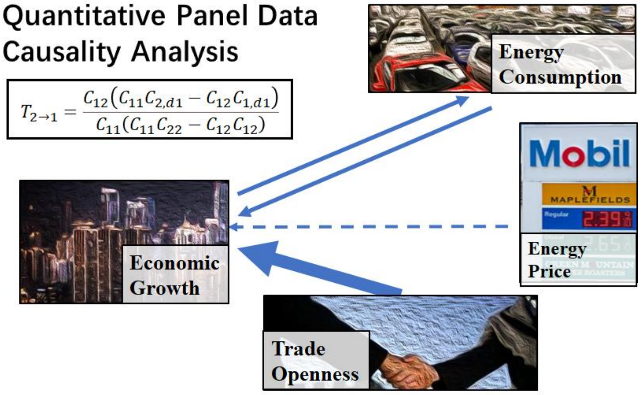

Fig. 5 Significant information flows among energy consumption, economic growth, trade openness and energy price. See the text for details.

The above inferred causal relations are evidenced by reports in the literature. First, the bidirectional causal relationship between energy consumption and economic growth has been discussed in many papers. Since the energy crisis in 1970s, many studies have confirmed the existence of such a causal relationship, e.g., [32]–[36], among others. Recently in some studies it is argued that no direct causal relationship between energy consumption and economic growth may exist [37]–[39]. Even this is true, most of such studies are based on the data from developed countries. For the 15 countries selected here, most are developing countries. The improvement of people's living standard is bound to the increase in energy consumption. Indeed, other studies based on the data from South Asia[40], [41], Southeast and East Asia [42], [43] all attest to this mutual causal relation.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3068273, IEEE Access

TABLE 2
THE ABSOLUTE VALUES OF INFORMATION FLOW (UNITS: NATS/YEAR) AND THE P-VALUES (IN PARENTHESES) AMONG THE ENERGY CONSUMPTION, ECONOMIC GROWTH, TRADE OPENNESS AND OIL PRICE.

Dependent variables	Source of causality			
	Energy consumption	Economic growth	Trade openness	Oil price
Energy consumption	/	**0.0102** **(0.0312)**	0.0011 (0.3142)	0.0000 (0.8801)
Economic growth	**0.0098** **(0.0116)**	/	**0.3142** **(0.0054)**	**0.0004** **(0.1490)**
Trade openness	0.0008 (0.6058)	0.0034 (0.5390)	/	0.0001 (0.8027)
Oil price	0.0000 (0.9240)	0.0001 (0.8482)	0.0001 (0.8628)	/

The values that are significant at 85% confidence level are in bold face.

For these 15 Asian countries, trade openness will not directly affect energy consumption; the converse does not hold, either. However, trade openness can affect the energy consumption by influence the economic growth. This is similar to the conclusion of Cole's [44], who found that trade liberalization promotes economic growth, which then boosts energy demand. It is noted that these 15 countries, especially those from East Asia and Southeast Asia, and India, have taken over a large portion of the manufacture from Europe and the United States since the 1980s, promoting economic growth and henceforth energy consumption through the globalized industrial chain. A slightly counter intuitive finding is that no direct causality between oil price and energy consumption is identified. But with the unidirectional causal link from oil price to energy consumption, oil price can exert impact on energy consumption. This does make more sense than a direct causality from oil price to energy consumption---Based on our observation, we would not drive more just because gasoline becomes cheaper. For oil importing/exporting countries, the rise in oil price is negatively/positively correlated with economic growth [45]. By influencing the economic growth, oil price may affect energy consumption to a certain extent. Sarwar et. al. [46] point out that fluctuations in oil price will affect economic growth, but electricity consumption can compensate for this effect to a certain extent. This is also the possible reason why oil price has only a weak impact on economic growth.

VI. CONCLUSION

Since it was found that information flow (IF) and causality are real physical notions and can be formulated on a rigorous footing (see [12]), many efforts have been made to put it to application to the important field of causal inference in data science. In this study, we generalized the method for time series, as established by Liang [14], to causal inference for homogeneous and i.i.d. panel data. The generalization is mathematically rigorous but straightforward, and the resulting formula bear the same form as that for time series, though the meanings of the symbols differ. We then proposed an algorithm, **Algorithm-IF**, for homogeneous and i.i.d. panel data causality analysis.

The algorithm has been validated with panel data sets from a linear stochastic model and a highly chaotic deterministic system. Three kinds of datasets, namely, time series, temporal style long sequences, and panel literature, have been generated and used for the validation. We found that in all these cases, the algorithm proves to be successful. Particularly, the performance with a touch-stone highly nonlinear problem proposed by Hahs and Pethel[28] turns out to be remarkably successful, though currently a linear assumption is made, in sharp contrast to the classical inference problem as discovered by Hahs and Pethel [28].

As a real-world application, we applied the algorithm to examine the causal relation among economic growth, energy consumption, trade openness, and energy price based on 15 Asian countries over the period 1980-2011. It is found that there are a strong bidirectional causality between economic growth and energy consumption, and a strong causality from import and export trade to economic growth.

Energy price does not have a direct impact on energy consumption, but it does exert a limited effect on the latter through influencing economic growth. These inferred causal relations are rather robust, and have been well justified by previous studies and observations.

Some issues remain. Recall the assumptions we have made in **Theorem III.1**, homogeneity and independence (and identical distribution). But a general panel dataset may be heterogeneous and may be subject to pervasive crosss-sectional dependence. For heterogeneous panel data, where some individuals may be causal while others may not be (e.g., [47]), more than one dynamical system should be taken into account in arriving at the information flow. For panel data with cross-sectional dependence, whereby all units in the same cross-section are correlated due to, for instance, the presence of common shocks and unobserved components that have been taken as part of the error ([48], [49]), the problem becomes more severe. These issues, among others, are to be investigated in future studies.

Acknowledgments. The constructive comments from the anonymous reviewers are appreciated. This work was partially supported by the National Science Foundation of China under Grant No. 41975064, and the 2015 Jiangsu Program for Innovation Research and Entrepreneurship Groups.

REFERENCES

[1] F. Xia, L. T. Yang, L. Wang, and A. Vinel, "Internet of things," *Int. J. Commun. Syst.*, vol. 25, no. 9, pp. 1101–1102, 2012, doi: https://doi.org/10.1002/dac.2417.

[2] H. U. Buhl, M. Röglinger, F. Moser, and J. Heidemann, "Big data," *WIRTSCHAFTSINFORMATIK*, vol. 55, no. 2, pp. 63–68, Apr. 2013, doi: 10.1007/s11576-013-0350-x.

[3] D. Che, M. Safran, and Z. Peng, "From big data to big data mining: challenges, issues, and opportunities," in *Database Systems for Advanced Applications*, Berlin, Heidelberg, 2013, pp. 1–15, doi: 10.1007/978-3-642-40270-8_1.

[4] M. Pesaran, *Time Series and Panel Data Econometrics*. 2015.

[5] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, 1969, doi: 10.2307/1912791.

[6] M. Paluš, A. Krakovská, J. Jakubík, and M. Chvosteková, "Causality, dynamical systems and the arrow of time," *Chaos*

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3068273, IEEE Access

Interdiscip. J. Nonlinear Sci., vol. 28, no. 7, p. 075307, Jul. 2018, doi: 10.1063/1.5019944.

[7] D. Holtz-Eakin, W. Newey, and H. S. Rosen, “Estimating vector autoregressions with panel data,” *Econometrica*, vol. 56, no. 6, pp. 1371–1395, 1988, doi: 10.2307/1913103.

[8] R. Hoffmann, C.-G. Lee, B. Ramasamy, and M. Yeung, “FDI and pollution: a granger causality test using panel data,” *J. Int. Dev.*, vol. 17, no. 3, pp. 311–317, 2005, doi: https://doi.org/10.1002/jid.1196.

[9] L. Kónya, “Exports and growth: Granger causality analysis on OECD countries with a panel data approach,” *Econ. Model.*, vol. 23, no. 6, pp. 978–992, Dec. 2006, doi: 10.1016/j.econmod.2006.04.008.

[10] S. Bedir and V. M. Yilmaz, “CO2 emissions and human development in OECD countries: granger causality analysis with a panel data approach,” *Eurasian Econ. Rev.*, vol. 6, no. 1, pp. 97–110, Apr. 2016, doi: 10.1007/s40822-015-0037-2.

[11] P. Gupta and A. Singh, “Causal nexus between foreign direct investment and economic growth: A study of BRICS nations using VECM and Granger causality test,” *J. Adv. Manag. Res.*, vol. 13, no. 2, pp. 179–202, Aug. 2016, doi: 10.1108/JAMR-04-2015-0028.

[12] X. S. Liang, “Information flow and causality as rigorous notions ab initio,” *Phys. Rev. E*, vol. 94, no. 5, p. 052201, Nov. 2016, doi: 10.1103/PhysRevE.94.052201.

[13] X. S. Liang and R. Kleeman, “Information transfer between dynamical system components,” *Phys. Rev. Lett.*, vol. 95, no. 24, p. 244101, Dec. 2005, doi: 10.1103/PhysRevLett.95.244101.

[14] X. S. Liang, “Unraveling the cause-effect relation between time series,” *Phys. Rev. E*, vol. 90, no. 5, p. 052150, Nov. 2014, doi: 10.1103/PhysRevE.90.052150.

[15] A. Stips, D. Macias, C. Coughlan, E. Garcia-Gorriz, and X. S. Liang, “On the causal structure between CO2 and global temperature,” *Sci. Rep.*, vol. 6, p. 21691, Feb. 2016, doi: 10.1038/srep21691.

[16] H. Xiao, F. Zhang, L. Miao, X. S. Liang, K. Wu, and R. Liu, “Long-term trends in Arctic surface temperature and potential causality over the last 100 years,” *Clim. Dyn.*, vol. 55, no. 5–6, pp. 1443–1456, Sep. 2020, doi: 10.1007/s00382-020-05330-2.

[17] C. Bai, R. Zhang, S. Bao, X. San Liang, and W. Guo, “Forecasting the tropical cyclone genesis over the Northwest Pacific through identifying the causal factors in cyclone–climate interactions,” *J. Atmospheric Ocean. Technol.*, vol. 35, no. 2, pp. 247–259, Feb. 2018, doi: 10.1175/JTECH-D-17-0109.1.

[18] G. Wang, C. Zhao, M. Zhang, Y. Zhang, M. Lin, and F. Qiao, “The causality from solar irradiation to ocean heat content detected via multi-scale Liang–Kleeman information flow,” *Sci. Rep.*, vol. 10, no. 1, Art. no. 1, Oct. 2020, doi: 10.1038/s41598-020-74331-2.

[19] H. Zhang, Z. Qiu, D. Sun, S. Wang, and Y. He, “Seasonal and interannual variability of satellite-derived Chlorophyll-a (2000–2012) in the Bohai Sea, China,” *Remote Sens.*, vol. 9, no. 6, p. 582, Jun. 2017, doi: 10.3390/rs9060582.

[20] D. F. Hagan, G. Wang, L. X. S., and H. A. J. Dolman, “A time-varying causality formalism based on the Liang–Kleeman information flow for analyzing directed interactions in nonstationary climate systems,” *J. Clim.*, vol. 32, no. 21, pp. 7521–7537, Nov. 2019, doi: 10.1175/JCLI-D-18-0881.1.

[21] X. S. Liang, “Normalizing the causality between time series,” *Phys. Rev. E*, vol. 92, no. 2, p. 022126, Aug. 2015, doi: 10.1103/PhysRevE.92.022126.

[22] D. T. Hristopulos, A. Babul, S. Babul, L. R. Brucar, and N. Virji-Babul, “Disrupted information flow in resting-state in adolescents with sports related concussion,” *Front. Hum. Neurosci.*, vol. 13, 2019, doi: 10.3389/fnhum.2019.00419.

[23] G. Schurz and A. Gebharter, “Causality as a theoretical concept: explanatory warrant and empirical content of the theory of causal nets,” *Synthese*, vol. 193, no. 4, pp. 1073–1103, Apr. 2016, doi: 10.1007/s11229-014-0630-z.

[24] X. S. Liang, “Information flow within stochastic dynamical systems,” *Phys. Rev. E*, vol. 78, no. 3, p. 031113, Sep. 2008, doi: 10.1103/PhysRevE.78.031113.

[25] X. S. Liang, “Causation and information flow with respect to relative entropy,” *Chaos Interdiscip. J. Nonlinear Sci.*, vol. 28, no. 7, p. 075311, Jul. 2018, doi: 10.1063/1.5010253.

[26] X. S. Liang, “The Liang-Kleeman information flow: theory and applications,” *Entropy*, vol. 15, no. 1, pp. 327–360, Jan. 2013, doi: 10.3390/e15010327.

[27] X. Lu, L. Su, and H. White, “Granger causality and structural causality in cross-section and panel data,” *Econom. Theory*, vol. 33, no. 2, pp. 263–291, Apr. 2017, doi: 10.1017/S0266466616000086.

[28] D. W. Hahs and S. D. Pethel, “Distinguishing Anticipation from Causality: Anticipatory Bias in the Estimation of Information Flow,” *Phys. Rev. Lett.*, vol. 107, no. 12, p. 128701, Sep. 2011, doi: 10.1103/PhysRevLett.107.128701.

[29] S. Nasreen and S. Anwar, “Causal relationship between trade openness, economic growth and energy consumption: A panel data analysis of Asian countries,” *Energy Policy*, vol. 69, pp. 82–91, Jun. 2014, doi: 10.1016/j.enpol.2014.02.009.

[30] F. Robert C., R. Inklaar, and M. P. Timmer, “The next generation of the penn world table,” *Am. Econ. Rev.*, vol. 105, no. 10, pp. 3150–3182, 2015, doi: 10.15141/S5Q94M.

[31] “BP statistical review of world energy 2013,” p. 48.

[32] J. Kraft and A. Kraft, “On the relationship between energy and GNP,” *J. Energy Dev.*, vol. 3, no. 2, pp. 401–403, 1978.

[33] A. T. Akarca and T. V. Long, “Energy and employment: a time-series analysis of the causal relationship,” *Resour. Energy*, vol. 2, no. 2–3, pp. 151–162, Oct. 1979, doi: 10.1016/0165-0572(79)90027-6.

[34] C. Nondo, M. Kahsai, and P. Schaeffer, “Energy consumption and economic growth: evidence from COMESA countries,” Regional Research Institute, West Virginia University, Working Paper 2010-01, 2010. Accessed: Jan. 24, 2021. [Online]. Available: https://ideas.repec.org/p/rri/wpaper/2010wp01.html.

[35] J. Baek and H. Kim, “Trade liberalization, economic growth, energy consumption and the environment: time series evidence from G-20 economies,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2318310, Nov. 2010. Accessed: Jan. 24, 2021. [Online]. Available: https://papers.ssrn.com/abstract=2318310.

[36] K. Saidi and S. Hammami, “Energy consumption and economic growth nexus: empirical evidence from Tunisia,” *Am. J. Energy Res.*, vol. 2, no. 4, pp. 81–89, Aug. 2014, doi: 10.12691/ajer-2-4-2.

[37] E. S. H. Yu and J. Choi, “The causal relationship between energy and GNP: An international comparison,” *J Energy Dev U. S.*, vol. 10:2, Jan. 1985.

[38] Z. Asghar, “Energy-GDP relationship: a causal analysis for the five countries of South Asia,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 1308260, Nov. 2008. Accessed: Jan. 27, 2021. [Online]. Available: https://papers.ssrn.com/abstract=1308260.

[39] A. Amirat and P. Bouri, “Energy and economic growth: the algerien case,” Jan. 2021.

[40] P. Mozumder and A. Marathe, “Causality relationship between electricity consumption and GDP in Bangladesh,” *Energy Policy*, vol. 35, no. 1, pp. 395–402, Jan. 2007, doi: 10.1016/j.enpol.2005.11.033.

[41] S. Noor and M. W. Siddiqi, “Energy consumption and economic growth in South Asian countries: a co-integrated panel analysis,” *Int. J. Energy Power Eng.*, vol. 4, no. 7, p. 6, 2010.

[42] H. A. Bekhet and N. Y. M. Yusop, “Assessing the relationship between oil prices, energy consumption and macroeconomic performance in Malaysia: co-integration and vector error correction model (VECM) approach,” *Int. Bus. Res.*, vol. 2, no. 3, Art. no. 3, Jun. 2009, doi: 10.5539/ibr.v2n3p152.

[43] T. N. Tran, T. T. Nguyen, V. C. Nguyen, and T. T. H. Vu, “Energy consumption, economic growth and trade balance in East Asia: a panel data approach,” *Int. J. Energy Econ. Policy*, vol. 10, no. 4, pp. 443–449, May 2020, doi: 10.32479/ijeep.9401.

[44] M. A. Cole, “Does trade liberalization increase national energy use?,” *Econ. Lett.*, vol. 92, no. 1, pp. 108–112, Jul. 2006, doi: 10.1016/j.econlet.2006.01.018.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

[45] L. Ghalayini, "The interaction between oil price and economic growth," *Middle East. Finance Econ.*, no. 13, Art. no. 13, 2011.

[46] S. Sarwar, W. Chen, and R. Waheed, "Electricity consumption, oil price and economic growth: Global perspective," *Renew. Sustain. Energy Rev.*, vol. 76, pp. 9–18, Sep. 2017, doi: 10.1016/j.rser.2017.03.063.

[47] E.-I. Dumitrescu and C. Hurlin, "Testing for Granger non-causality in heterogeneous panels," *Econ. Model.*, vol. 29, no. 4, pp. 1450–1460, Jul. 2012, doi: 10.1016/j.econmod.2012.02.014.

[48] R. E. De Hoyos and V. Sarafidis, "Testing for cross-sectional dependence in panel-data models," *Stata J.*, vol. 6, no. 4, pp. 482–496, Nov. 2006, doi: 10.1177/1536867X0600600403.

[49] M. H. Pesaran, "Estimation and inference in large heterogeneous panels with a multifactor error structure," *Econometrica*, vol. 74, no. 4, pp. 967–1012, 2006, doi: https://doi.org/10.1111/j.1468-0262.2006.00692.x.

Yineng Rong is a Ph.D. candidate at Nanjing University of Information Science and Technology. He is currently working on the development of causality-based machine learning algorithms with the Liang-Kleeman information flow, and their applications.

X. San Liang received his Ph.D. in Applied Mathematics from Harvard University, Massachusetts, USA. He has worked at the Second Institute of Oceanography of the State Oceanic Administration, Harvard University, Courant Institute, MIT, China Institute for Advanced Study, Central University of Finance and Economics, etc., and was a team leader of the Ninth Chinese National Antarctic Research Expedition. Presently he is Jiangsu Chair Professor at Nanjing University of Information Science and Technology, China. He is interested in a variety of interdisciplinary fields such as, not exclusively, complex systems, turbulence, multiscale modeling and simulation, predictability, causality analysis, information flow, manifold learning, atmosphere-ocean-climate science, etc.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/