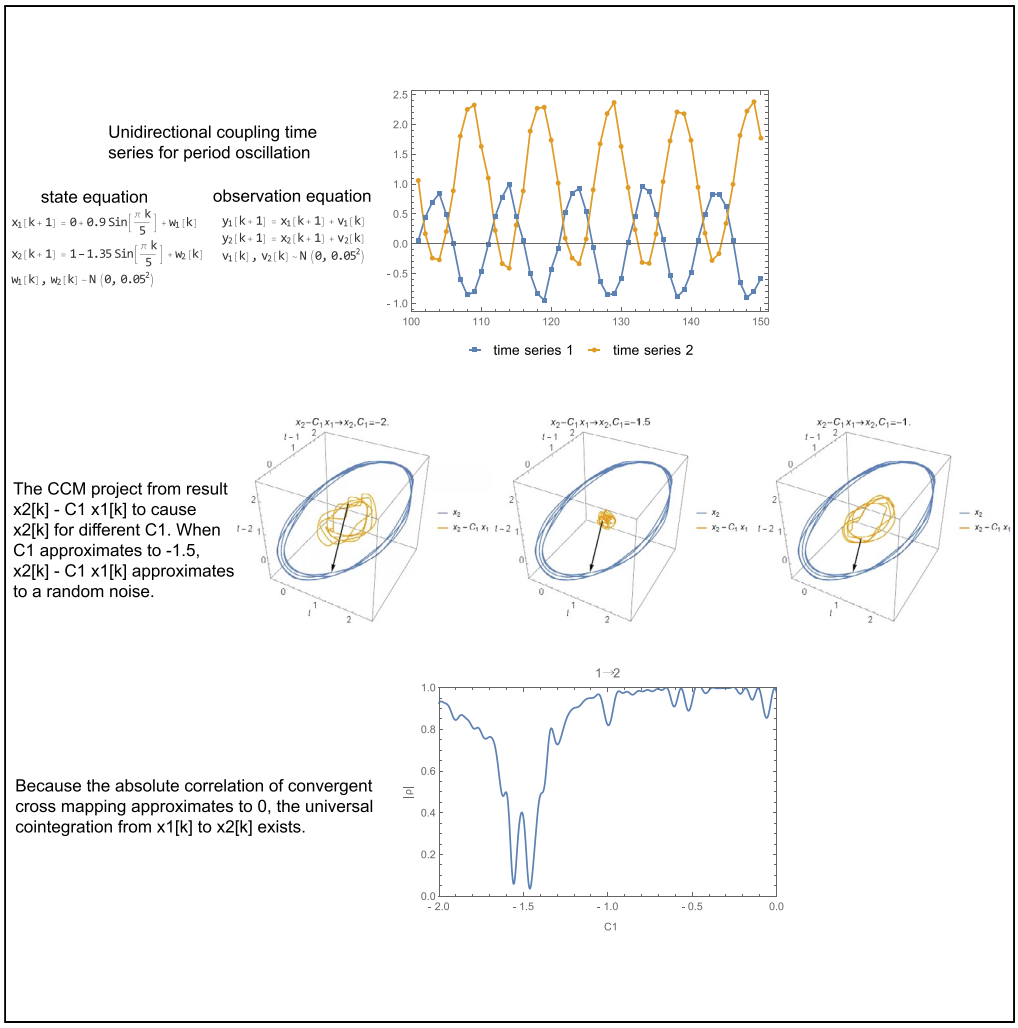
iScience

CellPress

Article

Universal Cointegration and Its Applications



Chengyi Tu, Ying Fan, Jianing Fan

chengyitu@berkeley.edu

HIGHLIGHTS

Specifically returning to the essence of the cointegration definition

Introducing a method to explore the universal cointegration

Demonstrating the method using controlled mathematical model examples

Applying the method to three real-world examples

Tu et al., iScience *19*, 986–995
September 27, 2019
https://doi.org/10.1016/j.isci.2019.08.048

iScience

CellPress

Article

Universal Cointegration and Its Applications

Chengyi Tu,[1,2,5,*] Ying Fan,[3] and Jianing Fan[4]

SUMMARY

Cointegration focuses on whether the long-term linear relationship between two or more time series is stationary even if this linear relationship does not exist or is not strong for the short term. Identifying the potential cointegration is important for economics, ecology, meteorology, neuroscience, and much more. Classic methods only considered or restricted in cointegration where the order of integration of all time series is 1. We introduce a method based on searching the vector to minimize the absolute correlation of convergent cross-mapping that can explore the universal cointegration and its extent. The proposed method can be applied to time series whose order of integration is not 1, cases that are not covered by classic cointegration. The proposed method is first illustrated and validated through time series generated by mathematical models in which the underlying relationships are known and then applied to three real-world examples.

INTRODUCTION

Identifying the potential cointegration among time series is a challenging and open problem (Johansen, 1988; Engle and Granger, 1991; Hamilton, 1994). Cointegration focuses on whether the long-term linear relationship between two or more time series is stationary even if this linear relationship does not exist or is not strong for the short term. For example, many models can be constructed based on well-established physical principles (modeling by mechanism). However, simulations of time series generated from these models may vary substantially. If a given model has the ability to capture the empirical observations, then the simulated time series of this model is expected to have a cointegrating relationship with the observed empirical time series. Therefore, for many candidate models, through the cointegrating relationship between the observed and simulated time series, one can determine which model is the most appropriate, i.e., the one that has the strongest cointegration. Such types of problems are ubiquitous in natural and social sciences (Kammerdiner and Pardalos, 2010; Kristoufek, 2013; Klee et al., 1987; Pedroni, 2001; Chiu and Wong, 2011; Ma and Zhu, 2019) and are difficult to discuss in complex nonlinear systems (Robinson and Hualde, 2003; Tu, 2014; Yang et al., 2014).

Cointegration is common in even the simplest nonlinear systems, such as shown in Figure 1, where time series are generated by the following set of state $\{x_1, x_2\}$ and observation $\{y_1, y_2\}$ equations:

$$\begin{cases} x_1[k+1] = \mu_1 + A_{11}G_1(x_1[k]) + A_{12}G_2(x_2[k]) + w_1[k] \\ x_2[k+1] = \mu_2 + A_{21}G_1(x_1[k]) + A_{22}G_2(x_2[k]) + w_2[k] \end{cases} \quad \text{(Equation 1)}$$

where $w_1[k], w_2[k]$ are independent normal distributions. If the matrix rank of A is 1, a constant α will exist such that $\frac{A_{21}}{A_{11}} = \frac{A_{22}}{A_{12}} = \alpha$, $\frac{A_{21}}{A_{11}} = \alpha, A_{22} = A_{12} = 0$ or $A_{21} = A_{11} = 0, \frac{A_{22}}{A_{12}} = \alpha$. Therefore, the above equation reads as

$$\begin{cases} x_1[k+1] = \mu_1 + A_{11}G_1(x_1[k]) + A_{12}G_2(x_2[k]) + w_1[k] \\ x_2[k+1] = \mu_2 + \alpha A_{11}G_1(x_1[k]) + \alpha A_{12}G_2(x_2[k]) + w_2[k] \end{cases} \quad \text{(Equation 2)}$$

The relationship between $x_1[k+1]$ and $x_2[k+1]$ is $x_2[k+1] = \alpha x_1[k+1] + (\mu_2 - \alpha\mu_1) + (w_2[k] - \alpha w_1[k])$. At each step, the observation $\mathbf{y}[k+1]$ of the true state $\mathbf{x}[k+1]$ is

$$\begin{cases} y_1[k+1] = x_1[k+1] + v_1[k+1] \\ y_2[k+1] = x_2[k+1] + v_2[k+1] \end{cases} \quad \text{(Equation 3)}$$

[1]School of Ecology and Environmental Science, Yunnan University, Kunming, China

[2]Department of Environmental Science, Policy, and Management, University of California, Berkeley, Berkeley, CA, USA

[3]College of Geography and Environment, Shandong Normal University, Jinan, China

[4]School of Economics and Management, Beijing Information Science & Technology University, Beijing, China

[5]Lead Contact

*Correspondence: chengyitu@berkeley.edu

https://doi.org/10.1016/j.isci.2019.08.048

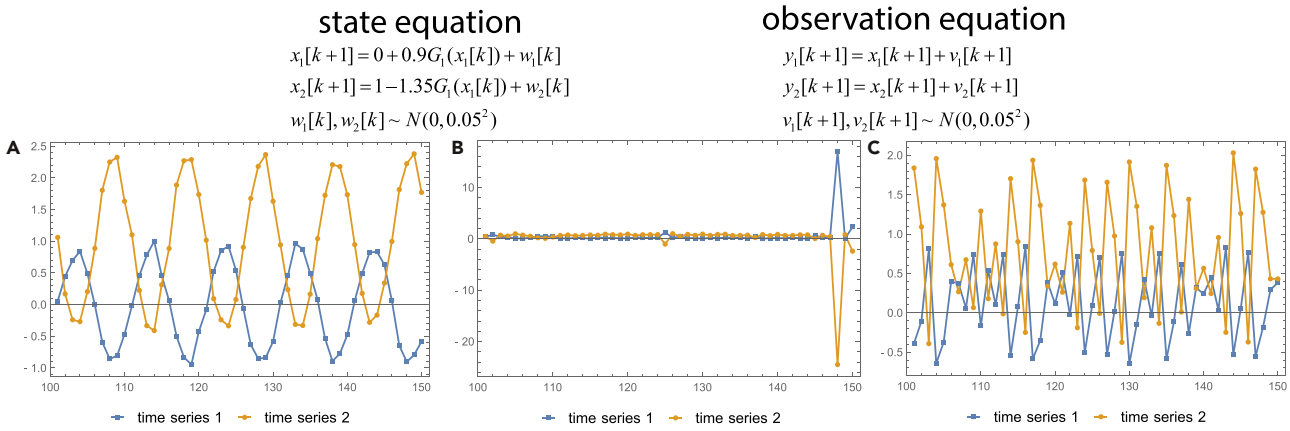


Figure 1. Unidirectional Coupling Time Series

(A) Period oscillation where $G_1(x_1[k]) = \mathrm{Sin}\left(\frac{\pi k}{5}\right)$.

(B) Extreme value where $G_1(x_1[k])$ is drawn from a Pareto distribution with a minimum value parameter of 0.1 and a shape parameter of 1.25.

(C) Chaotic map where $G_1(x_1[k]) = 1 - 2|x_1[k]|$. The initial value is drawn from a uniform distribution between 0 and 1. Steps from 101 to 150 are adopted. Although a short-term deviation between the two time series exists, the long-term relationship is stationary according to the equations.

where $v_1[k+1], v_2[k+1]$ are independent normal distributions. Therefore, the following equation holds:

$$y_2[k+1] = \alpha y_1[k+1] + (\mu_2 - \alpha\mu_1) + (w_2[k] - \alpha w_1[k] + v_1[k+1] + v_1[k+1]) \quad \text{(Equation 4)}$$

where $\mu_2 - \alpha\mu_1$ is a constant and $w_2[k] - \alpha w_1[k] + v_1[k+1] + v_2[k+1]$ is a normal distribution. Although $y_2[k+1]$ deviates from $\alpha y_1[k+1] + (\mu_2 - \alpha\mu_1)$ owing to the noise for short time windows, their difference is stationary if we consider sufficiently long time intervals.

The traditional statistical method for identifying the linear relationship between two time series is regression. The most well-known disadvantage of this method is spurious regression (Granger and Newbold, 1974), which is misleading statistical evidence of a linear relationship between independent nonstationary variables (see Transparent Methods, section Spurious regression). Therefore, applying a regression model to identify long-term stationary relationships is not reliable and valid (see Figure S1 and Table S1), also considering that the empirical time series are contaminated by diverse factors. Engle and Granger realized this critical problem early on and proposed the concept of cointegration (Engle and Granger, 1987).

Classic cointegration, such as the Engle-Granger cointegration test (Engle and Granger, 1987) and Johansen cointegration test (Johansen, 1995), only considers or restricts in cointegration $CI(1,1)$ (Hamilton, 1994), i.e., (1) all time series $\mathbf{x}[k]$ are nonstationary, and the minimum number of differences required to obtain the stationary series is 1, and (2) a vector $\mathbf{C} \neq 0$ exists such that $\mathbf{C}^T\mathbf{x}[k]$ is stationary (see Transparent Methods, section Background of cointegration). The Engle-Granger cointegration test, based on ordinary least squares, seeks the linear combination that has the minimum variance (see Table S2). The Johansen cointegration test, based on the maximum likelihood estimator of the so-called reduced rank model, seeks the linear combination that is the most stationary (see Table S3). The prerequisite of these approaches is that each given time series is $I(1)$, namely, it is nonstationary, and the minimum number of differences required to obtain a stationary series is 1. Otherwise, preprocessing such as difference or log-transform will be adopted, for example, asset returns in financial time series (Tsay, 2005) and phase estimation of neurophysiological signals (Dahlhaus et al., 2017). However, these preprocessings cannot guarantee that all processed time series are all $I(1)$ together. Even if so, the following cointegration test is applied to the processed time series rather than the original time series. Therefore, the result is not reliable and clear.

However, even if the prerequisite that all time series are all $I(1)$ is not satisfied, these methods can still be run mechanically (see Transparent Methods, section Application of classic cointegration and its misuse). For example, the time series generated by Figure 1 cannot be applied to classic cointegration methods (see Tables S4 and S5). In the three cases, the long-term linear relationships exist, but all time series are not $I(1)$, and classic cointegration methods are violated completely. Unfortunately, the time series of

most empirical datasets, such as economics, ecology, meteorology, and neuroscience, often cannot satisfy this prerequisite.

In addition to this drawback, classic cointegration analysis also has two other drawbacks (Hamilton, 1994; Dahlhaus et al., 2018). First, it is difficult or even impossible to handle time series where any preprocessing cannot obtain $I(1)$ together. Second, it is invalid for (complete) synchronization coupling (Pikovsky et al., 2003) composed of time series from a chaos system (see Transparent Methods, section Background of cointegration). The synchronization guarantees that all time series are the same or at least similar, so their linear combination is stationary (Pikovsky et al., 2003). Therefore, synchronization and cointegration describe the same problem: dynamic fluctuation around the equilibrium (Dahlhaus et al., 2018).

In this work, we examine a method that specifically returns to the essence of the cointegration definition, only considering whether the long-term linear relationship between two or more time series exists and its extent. We demonstrate the principles of our framework using controlled mathematical model examples, showing that the method can successfully identify the cointegration that is not covered by classic cointegration but important in practice. The method is particularly suitable for identifying the synchronization naturally and heuristically. Finally, we apply the method to (1) check the relationship between models and observations of global warming, (2) identify the possible synchronization in electroencephalographic signals, and (3) determine the possible leadership of Bitcoin in the cryptocurrency market.

Our method is not in competition with the classic cointegration methods, such as the Engle-Granger cointegration test and Johansen cointegration test; rather, it is specifically aimed at a class of systems not covered by classic cointegration (see Transparent Methods, section Classic cointegration analysis). The classic cointegration cannot apply to all cases in this work because its prerequisite is violated, even if it can occasionally determine the existing cointegrating relationship (see Transparent Methods, section Application of classic cointegration and its misuse and Tables S6 and S7).

RESULTS

Mathematical Model Examples

When the system size is two, the possible cointegrating vector is $\mathbf{C} = \{C_1, -1\}$. Three cases to construct the cointegrating relationship are discussed: unidirectional coupling, bidirectional coupling, and synchronization coupling (see Transparent Methods, section Mathematical model to generate the time series data).

Unidirectional Coupling

Considering the case where $\mathbf{A} = \begin{bmatrix} 0.9 & 0 \\ -1.35 & 0 \end{bmatrix}$ (see Figure 1); if we assume that time series $x_2[k]$ is dependent and $x_1[k]$ is independent, the minimum of the absolute correlation of CCM $|\rho| \to 0$, and the argument $C_1 \to \frac{A_{21}}{A_{11}} = \frac{-1.35}{0.9} = -\frac{3}{2}$ (see the first row of Figure 2). To better understand the details, Video S1 shows the CCM projection from $x_2[k] - C_1 x_1[k]$ to $x_2[k]$ with different C_1 for the case in Figure 1A, related to Figure 1. When $C_1 \to -\frac{3}{2}$, $x_2[k] - C_1 x_1[k]$ approximates to a random noise. Therefore, the universal cointegration from $x_1[k]$ to $x_2[k]$ exists, and its extent is large. If we assume that $x_1[k]$ is dependent and $x_2[k]$ is independent, then the universal cointegration from $x_2[k]$ to $x_1[k]$ also exists, and the argument $C_1 \to \frac{A_{11}}{A_{21}} = \frac{0.9}{-1.35} = -\frac{2}{3}$ (see the second row of Figure 2). We display the exact minimum and its argument value in Table S8 and the summary of 50 different realizations in Figure 3A.

Bidirectional Coupling

Considering the case where $\mathbf{A} = \begin{bmatrix} 1/3 & 2/3 \\ 2/9 & 4/9 \end{bmatrix}$ (see Figure S3), if we assume that time series $x_2[k]$ is dependent and $x_1[k]$ is independent, the minimum of the absolute correlation of CCM $|\rho| \to 0$ and the argument $C_1 \to \frac{A_{21}}{A_{11}} = \frac{A_{22}}{A_{12}} = \frac{2}{3}$ (see the first row of Figure S4). Therefore, the universal cointegration from $x_1[k]$ to $x_2[k]$ exists, and its extent is large. If we assume that $x_1[k]$ is dependent and $x_2[k]$ is independent, the universal cointegration from $x_2[k]$ to $x_1[k]$ also exists, and the argument $C_1 \to \frac{A_{11}}{A_{21}} = \frac{A_{12}}{A_{22}} = \frac{3}{2}$ (see the second row of

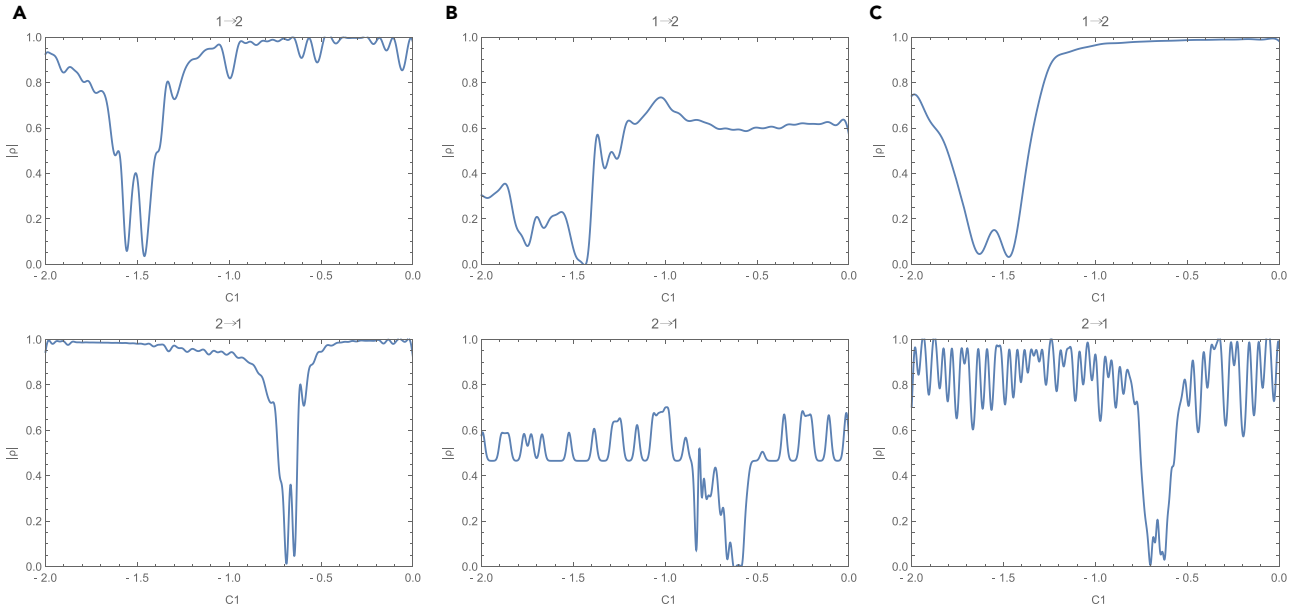


Figure 2. The Acquisition Functions of Bayesian Minimization for the Three Cases of Unidirectional Coupling Time Series

(A) Period oscillation.

(B) Extreme value.

(C) Chaotic map. The first row checks the universal cointegration from time series 1 to 2, and the second row checks the opposite direction.

Figure S4). We display the exact minimum and its argument value in Table S9 and the summary of 50 different realizations in Figure 3B.

Synchronization Coupling

Considering the case $\mathbf{A}=\begin{bmatrix}1-\epsilon & \epsilon\\ \epsilon & 1-\epsilon\end{bmatrix}$ and the coupling dynamics $G_1(x_1[k]) = G_2(x_2[k])$, if ϵ is or approximates to 0.5, then the two time series are always similar irrespective of the special function or even chaos function (see Figure S5). Even if noise exists in synchronized processes, their difference should remain relatively small, i.e., the differences follow a stationary process. Therefore, the minimum of the absolute correlation of CCM $|\rho|\to 0$, and the argument $C_1 \approx 1$ (see Figure S6). We display the exact minimum and its argument value in Table S10 and the summary of 50 different realizations in Figure 3C.

System Size Is Larger than Two

When the system size is larger than two, if we still restrict ourselves to discussing the possible universal cointegration between each pair (see Figures S7A–S7C), i.e., the size of the cointegrating vector is two, then the process is the same as before (see Figures S8A–S8C). If we extend to discussing the mixing coupling where the possible universal cointegration is among three or more time series, our method is still valid. Considering a simple example of mixing coupling where the adjacency matrix $\mathbf{A}=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 2/3 & 1/3 & 0\end{bmatrix}$ (see Transparent Methods, section Mathematical model to generate the time series data), $x_3[k] = \frac{2}{3}x_1[k] + \frac{1}{3}x_2[k] + \epsilon[k]$, i.e., $x_3[k]$ is dependent and $x_1[k]$ and $x_2[k]$ are both independent (see Figure S7D). Therefore, we search the vector $\mathbf{C}=\{C_1, C_2, -1\}$ to minimize the absolute correlation of CCM $|\rho|$ between cause $x_3[k]$ and result $x_3[k] - C_1x_1[k] - C_2x_2[k]$ by Bayesian minimization. If the minimum approximates to 0, i.e., $|\rho|\to 0$, the universal cointegration exists, and $\{C_1, C_2, -1\}$ is the desired cointegrating vector (see Figure S8D). The deeper the color is, the smaller is the absolute correlation of CCM. The sole deep color area indicates that there is only one cointegrating relationship among these time series.

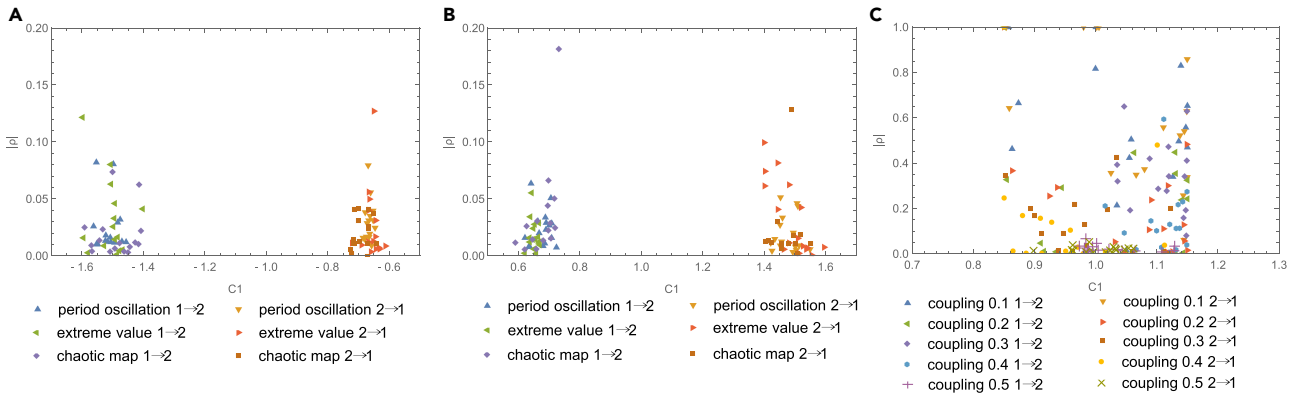


Figure 3. The Minimum and Its Argument for Different Cases

(A) Unidirectional coupling including three subcases.

(B) Bidirectional coupling including three subcases.

(C) Synchronization coupling including five coupling strengths. For each subcase, we run 50 realizations, and each market point represents the minimum and its argument of one realization.

Real-World Examples

Real-world data are always contaminated by diverse factors that can be reduced to observation noise and process noise. These noises introduce many short-term deviations but cannot force the disappearance of the long-term stationary relationship. Therefore, noises are a benefit rather than a curse for universal cointegration. Here, we will apply it to three simple but frontier examples.

Checking the Relationship between Models and Observations of Global Warming

Climate is the response to linkages and couplings between the atmosphere, the hydrosphere, the biosphere, the cryosphere, and the geosphere (Houghton et al., 2001; Beniston et al., 1997). Climate models based on well-established physical principles have been demonstrated not only to successfully reproduce observed features of climate changes but also to predict future changes (Allen and Tett, 1999; Zhang et al., 2007). However, the simulation results, particularly the patterns of processes and phenomena, may substantially vary from different models (Turasie, 2012; Kaufmann et al., 2011). If a climate model has the ability to capture the real system, then its outputs are expected to have a cointegrating relationship with the observed climate (Turasie, 2012). Do the outputs of climate models cointegrate with the observed climate change? If so, what is the best proxy comparing the performances of different models?

Our first real-world example is applying universal cointegration to compare the global near-surface temperature of the historical observations and simulations from 48 models in the CMIP5 archive (see Figure 4A). Some studies used the regression method or classic cointegration analysis (Turasie, 2012). As indicated in our previous discussions, these methods have some fatal drawbacks, and our method does not need to determine the order of integration. If the model time series $x_m[k]$ cointegrates with the observation time series $x_o[k]$, the minimum of the absolute correlation of CCM $|\rho| \to 0$ and $x_o[k] = C_1 x_m[k] + \alpha + \epsilon[t]$ holds. Furthermore, if the model time series is a good proxy of the observation time series, i.e., the observation time series is just the model time series with noise, then the cointegrating parameter $\alpha \to 0$ and $C_1 \to 1$. Therefore, we define the Euclidean distance in the three-dimensional space from $\{|\rho|, C_1, \alpha\}$ to $\{0, 1, 0\}$, i.e., $\sqrt{(\rho)^2 + (C_1 - 1)^2 + (\alpha)^2}$, as the criterion to check whether the model is a good proxy.

Most model time series have a cointegrating relationship with the observation time series due to $|\rho| \to 0$ (see Figure 4B), but the cointegrating parameters $\alpha < 0$, $C_1 > 1$. Therefore, most models successfully capture the real climate system, but they are always higher than the observation, i.e., these models overestimate the global warming. In particular, as time passes, the overestimated value will increase. We display the exact values $\{|\rho|, C_1, \alpha\}$ of each model in Table S11. Here, we just propose some warnings about the climate models rather than negating their significance (Oreskes, 2004).

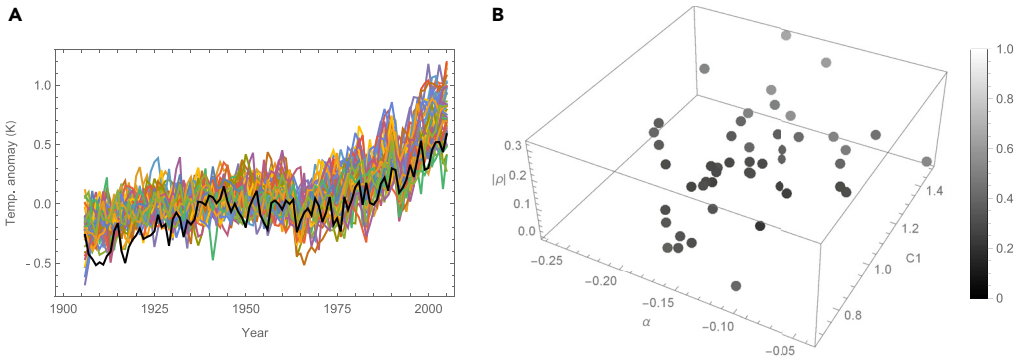


Figure 4. The Relationship between Models and Observations of Global Warming

(A) The observation time series and 48 model time series of temperature anomalies from 1906 to 2005. The black lines are observation time series, and the colored lines are model time series.

(B) The three-dimensional space $\{|\rho|, C_1, \alpha\}$. Each point represents a model time series. The darker the point is, the smaller is the distance to the critical point $\{0, 1, 0\}$.

Identifying the Possible Synchronization in Electroencephalographic Signals

The synchronization phenomena in neuroscience have been increasingly regarded as essential for the functional coupling of different brain regions (Varela et al., 2001; Fries et al., 2001; Izhikevich, 2007), and pathological synchronization has been regarded as a main mechanism responsible for an epileptic seizure (Traub and Wong, 1982; Iasemidis, 2003; Mormann et al., 2003). Many synchronization measures are proposed, namely, nonlinear interdependence (Le Van Quyen et al., 1998), mutual information (Jeong et al., 2001), cross-correlation (Chandaka et al., 2009), and coherence function (Shaw, 1981). Because both synchronization and universal cointegration describe dynamic fluctuations around the equilibrium, universal cointegration can determine the synchronization and its extent.

The potential synchronization between the left and right hemisphere rat electroencephalographic (EEG) channels is hard to guess beforehand from the raw data, and whether universal cointegration can provide a relevant contribution to the study of synchronization in EEG and whether it can disclose information are difficult to obtain by visual inspection. If the synchronization between the left and right time series is strong, the cointegrating relationship $x_{r/l}[k] = C_1 x_{l/r}[k] + \alpha + \epsilon[t]$ will hold, where $x_{l/r}[k]$ is left/right time series. We analyze the potential synchronization between the left and right EEG channels in three different cases.

Although the possible synchronization in each case is difficult to determine by visual inspection (see Figures 5A, S9A, S9C, and S9E), checking the universal cointegration and comparing the result is simple. We find that the synchronization extent of these cases is 1>3>2 in general. In case 1, the absolute correlations of universal cointegration from right to left and from left to right are both small; thus, its synchronization is always strong (see Figure 5B). The regions of dark color in case 2 display the disappearance of order. For different cell sizes, such as 50, 250, and 500, the results are similar (see Figures S9B, S9D, and S9F). Although we do not have objective means for claiming that the difference between the synchronization of the time series is significant, the quantification of synchronization between different time series can complement the conventional visual analysis and can even be of clinical value (Quiroga et al., 2002). The method is so simple, straightforward, and fast that it is very easy to be adopted as an online implementation. Additionally, the method should not be restricted to EEG data and can be valuable for studying the synchronization of other time series.

Determining the Possible Leadership of Bitcoin in the Cryptocurrency Market

In recent years, a new type of financial asset, cryptocurrency, has been introduced, and it is emerging as a new topic in empirical economic studies (Donier and Bouchaud, 2015; Gatfaoui et al., 2017; Tu et al., 2018). Notably, its market capitalization is approximately 300 billion dollars, and it is traded with many of the main national currencies with daily trading of more than 10 billion USD. The seminal and most popular cryptocurrency is Bitcoin (BTC), occupying half of the market capitalization and trade volume in the whole cryptocurrency market (ElBahrawy et al., 2017). The last half year has witnessed an unusual rise and fall in the price of cryptocurrency, emerging as an asset bubble (Bariviera, 2017; Blau, 2017); for example, the price

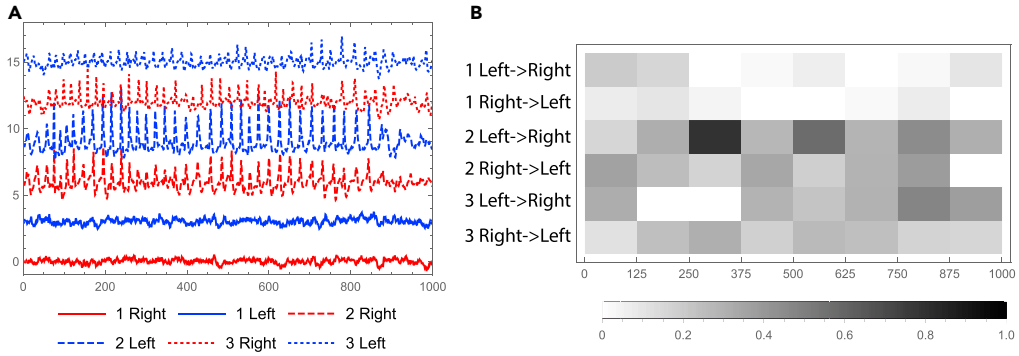


Figure 5. The Possible Synchronization in Electroencephalographic Signals

(A) Three cases of rat EEG signals/time series from right and left cortical intracranial electrodes. For better visualization, all time series are plotted with an offset step of 3.

(B) The minimum of the absolute correlation of CCM $|\rho|$ of each case. Each row represents a possible cointegration, and each cell represents its value calculated by 125 steps.

of BTC increased from approximately 4,000 dollars in early October 2017 to approximately 20,000 dollars at its peak in December 2017. Because there is no compelling way to assess the fundamentals of cryptocurrency price, the tremendous price fluctuation is elusory and instantaneous (Urquhart, 2017; Ametrano, 2016). Despite the theoretical and economic interest in the cryptocurrency market, a comprehensive analysis of whether BTC is the factual leader is still lacking. The classic cointegration analysis is not a good choice to answer this question because it is difficult or unrealistic to guarantee the prerequisite, i.e., all time series are $I(1)$ together even if after preprocessing of econometrics. As in the previous discussions, our universal cointegration is appropriate for addressing this question naturally and heuristically.

Our last real-world example focuses on whether BTC is the leader of other cryptocurrencies and its extent. If BTC is the leader, the cointegrating relationship from BTC to another cryptocurrency $x_c[k] = C_1 x_{BTC}[k] + \alpha + \epsilon[t]$ will hold, where $x_{BTC}[k]$ is the price time series of BTC and $x_c[k]$ is another cryptocurrency. The smaller the absolute correlation of CCM $|\rho|$ is, the stronger the leadership of BTC. Here, we consider the evolution of universal cointegration from BTC to ETH, XRP, LTC, EOS, ADA, and XLM between April 1, 2017 and March 31, 2018.

In general, the universal cointegration of all cryptocurrencies presents similar behavior (see Figure 6B). Before February 2018, the absolute correlation of each cointegration is approximately 0.1 with a small but saltatory fluctuation, i.e., the leadership of BTC weakly exists. Then, a cliffy peak with a width of a half month emerges in the middle of February 2018, and the peak value is larger than 0.6, i.e., the leadership of BTC disappears suddenly and extremely. After the peak, the value falls steeply and approximates to 0 at the beginning of March, i.e., the leadership of BTC strongly exists. Finally, it increases progressively, as does the leadership of BTC. Because the length of the time window is 90, we left shift 45 to discuss the relationship between the leadership and the price of BTC (see Figure 6A). In the corresponding region of the first region, from the beginning of October 2017 to the middle of December 2017, the price of BTC increases step by step. In the region from the middle of December 2017 to the end of December 2017, corresponding to the second region, the price of BTC maintains a high fluctuation with a relatively small shock. Then, from the beginning of January 2018 to the end of January 2018, the price of BTC undergoes a steep decrease. Finally, from the beginning of February 2018 to the middle of February 2018, the price of BTC increases gradually. Because BTC occupies half of the whole cryptocurrency market, its steep fall would lead to market jitters and then market collapse. Therefore, the behaviors of other cryptocurrencies are similar to those of BTC. Finally, the leadership of BTC strongly exists. If the price of BTC maintains a high fluctuation with a relatively small shock, hot money in the cryptocurrency market does not have an effective investment opportunity. Therefore, they will transfer to other cryptocurrencies. At last, the leadership of BTC disappears. In other cases, different cryptocurrencies present not only the behavior of BTC but also individual behavior. Therefore, the leadership of BTC may exist, but its extent is not enough. Other time windows will obtain similar results (see Figure S10).

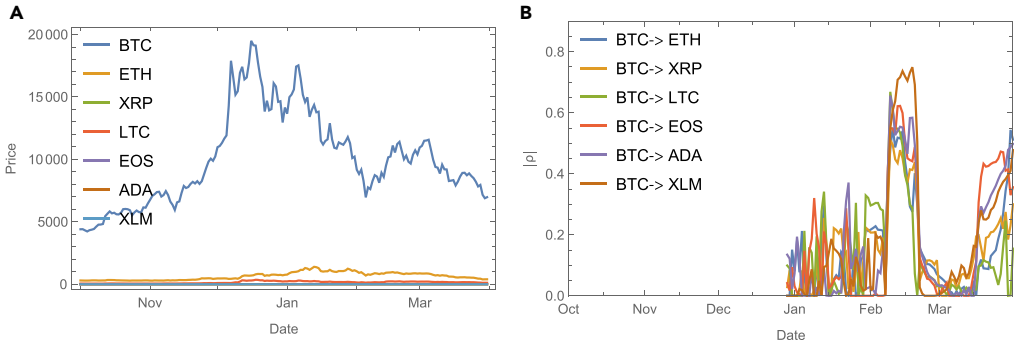


Figure 6. The Possible Leadership of Bitcoin in the Cryptocurrency Market

(A) The price time series of seven typical cryptocurrencies, including Bitcoin (BTC), Ethereum (ETH), Ripple (XRP), Litecoin (LTC), EOS (EOS), Cardano (ADA), and Stellar (XLM), from April 1, 2017, to March 31, 2018.

(B) The absolute correlation of CCM from BTC to other cryptocurrencies. The time window is 90, approximately half of the length of time series.

DISCUSSION

The classic cointegration analysis, such as Engle-Granger cointegration and Johansen cointegration, only considers or restricts in $CI(1,1)$. These methods assume that all time series are $I(1)$ and a vector exists such that the linear combination is $I(0)$; otherwise, preprocessing will be adopted. However, even if the prerequisite that all time series are all $I(1)$ is not satisfied, these methods can still be run mechanically. Additionally, these preprocessings cannot guarantee that all processed time series are all $I(1)$ together. Even if so, the cointegration exists in the processed time series rather than the original time series. To address this, we introduce a method based on searching the vector to minimize the absolute correlation of CCM that can explore the universal cointegration and its extent. Our method is not in competition with the classic cointegration analysis; rather, it is specifically aimed at a class of systems not covered by classic cointegration analysis. Thus, it is not surprising that applying classic cointegration analysis to some examples fall outside of its range but covered by universal cointegration is largely uncertain and confusing (see Transparent Methods, section Application of classic cointegration and its misuse).

Many methods to construct complex networks from empirical data have been proposed, but we suggest that universal cointegration inferred from time series of each pair of nodes provides a complement to the conventional method. The ability to resolve a cointegration network from their dynamical behavior sheds new light on the underlying mechanisms and driving forces, particularly when it is important to know the importance of nodes interacting as a group and need to be considered together. For example, in financial markets, accurate knowledge of the cointegration network not only can offer the potential opportunity of pairs trading but also deepens our understanding of financial collapse (Tu, 2014; Iori and Mantegna, 2018).

Limitations of the Study

Because CCM is restricted to discrete-time settings, universal cointegration cannot apply to continuous-time settings directly. The main problem is that a continuous-time signal cannot construct a shadow manifold. However, in practice, we can transform a continuous-time signal into a discrete-time signal by sampling. In the manuscript, the time series of period oscillation $\mathrm{Sin}\left(\frac{\pi k}{5}\right)$ (see Figures 1A, S3A, and S7) is sampled from a continuous-time signal $\mathrm{Sin}(t)$ at a sampling interval of $\pi/5$. Classic cointegration obtains the parameters directly, such as C_1, α. Although it requires little calculation time, it does not test other parameters that may obtain a stronger cointegration, i.e., a vector $\mathbf{C} = \{C_1, \cdots, C_{N-1}, -1\}$ exists such that $x_N[k] - C_1 x_1[k] - \cdots - C_{N-1} x_{N-1}[k]$ more approximates to a noise. The calculation time required for universal cointegration is often larger than that of classic cointegration, and it mainly depends on the maximum number of iterations of Bayesian minimization. When this value increases, Bayesian minimization will be tried more times to obtain a smaller $|\rho|$, then the cointegration will be stronger.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

DATA AND CODE AVAILABILITY

All data and code are available in the Mendeley Data website https://doi.org/10.17632/7d9rm7yjxd.1.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2019.08.048.

ACKNOWLEDGMENTS

C.T. acknowledges part of the financial support comes from Yunnan University project C176210103.

AUTHOR CONTRIBUTIONS

C.T. conceived of, designed, and supervised the study. C.T., Y.F., and J.F. analyzed the data and wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 19, 2019
Revised: July 22, 2019
Accepted: August 25, 2019
Published: September 27, 2019

REFERENCES

Allen, M.R., and Tett, S.F. (1999). Checking for model consistency in optimal fingerprinting. Clim. Dyn. *15*, 419–434.

Ametrano, F.M. (2016). Hayek money: the cryptocurrency price stability solution. Available at SSRN 2425270, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2425270.

Bariviera, A.F. (2017). The inefficiency of bitcoin revisited: a dynamic approach. Econ. Lett. *161*, 1–4.

Beniston, M., Diaz, H., and Bradley, R. (1997). Climatic change at high elevation sites: an overview. Climatic Change *36*, 233–251.

Blau, B.M. (2017). Price dynamics and speculative trading in bitcoin. Res. Int. Bus. Finance *41*, 493–499.

Chandaka, S., Chatterjee, A., and Munshi, S. (2009). Cross-correlation aided support vector machine classifier for classification of EEG signals. Expert Syst. Appl. *36*, 1329–1336.

Chiu, M.C., and Wong, H.Y. (2011). Mean–variance portfolio selection of cointegrated assets. J. Econ. Dyn. Control *35*, 1369–1385.

Dahlhaus, R., Dumont, T., Le Corff, S., and Neddermeyer, J.C. (2017). Statistical inference for oscillation processes. Statistics *51*, 61–83.

Dahlhaus, R., Kiss, I.Z., and Neddermeyer, J.C. (2018). On the relationship between the theory of cointegration and the theory of phase synchronization. Stat. Sci. *33*, 334–357.

Donier, J., and Bouchaud, J.P. (2015). Why do markets crash? bitcoin data offers unprecedented insights. PLoS One *10*, e0139356.

ElBahrawy, A., Alessandretti, L., Kandler, A., Pastor-Satorras, R., and Baronchelli, A. (2017). Evolutionary dynamics of the cryptocurrency market. R. Soc. Open Sci. *4*, 170623.

Engle, R.F., and Granger, C.W. (1987). Co-integration and error correction: representation, estimation, and testing. Econometrica *55*, 251–276.

Engle, R., and Granger, C. (1991). Long-run Economic Relationships: Readings in Cointegration (Oxford University Press).

Fries, P., Reynolds, J.H., Rorie, A.E., and Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. Science *291*, 1560–1563.

Gatfaoui, H., Nagot, I., and De Peretti, P. (2017). Are critical slowing down indicators useful to detect financial crises? In Systemic Risk Tomography, R. Savona, M. Billio, and L. Pelizzon, eds. (Elsevier), pp. 73–93.

Granger, C.W., and Newbold, P. (1974). Spurious regressions in econometrics. J. Econom. *2*, 111–120.

Hamilton, J.D. (1994). Time Series Analysis (Princeton University Press).

Houghton, J.T., Ding, Y., Griggs, D.J., Noguer, M., van der Linden, P.J., Dai, X., Maskell, K., and Johnson, C. (2001). Climate Change 2001: The Scientific Basis (The Press Syndicate of the University of Cambridge).

Iasemidis, L.D. (2003). Epileptic seizure prediction and control. IEEE Trans. Biomed. Eng. *50*, 549–558.

Iori, G., and Mantegna, R.N. (2018). Empirical analyses of networks in finance. In Handbook of Computational Economics, *volume 4*, C. Hommes and B. LeBaron, eds., Handbook of Computational Economics (Elsevier), pp. 637–685.

Izhikevich, E.M. (2007). Dynamical Systems in Neuroscience (MIT press).

Jeong, J., Gore, J.C., and Peterson, B.S. (2001). Mutual information analysis of the EEG in patients with Alzheimer's disease. Clin. Neurophysiol. *112*, 827–835.

Johansen, S. (1988). Statistical analysis of cointegration vectors. J. Econ. Dyn. Control *12*, 231–254.

Johansen, S. (1995). Likelihood-based Inference in Cointegrated Vector Autoregressive Models (Oxford University Press on Demand).

Kammerdiner, A.R., and Pardalos, P.M. (2010). Analysis of multichannel EEG recordings based on generalized phase synchronization and cointegrated var. In Computational Neuroscience, W. Chaovalitwongse, P.M. Pardalos, and P. Xanthopoulos, eds. (Springer), pp. 317–339.

Kaufmann, R.K., Kauppi, H., Mann, M.L., and Stock, J.H. (2011). Reconciling anthropogenic climate change with observed temperature 1998–2008. Proc. Natl. Acad. Sci. U S A *108*, 11790–11793.

Klee, H., Horsch, R., and Rogers, S. (1987). Agrobacterium-mediated plant transformation and its further applications to plant biology. Annu. Rev. Plant Physiol. *38*, 467–486.

Kristoufek, L. (2013). Bitcoin meets google trends and wikipedia: quantifying the relationship between phenomena of the internet era. Sci. Rep. *3*, 3415.

Le Van Quyen, M., Adam, C., Baulac, M., Martinerie, J., and Varela, F.J. (1998). Nonlinear

interdependencies of EEG signals in human intracranially recorded temporal lobe seizures. Brain Res. *792*, 24–40.

Ma, G., and Zhu, S.P. (2019). Optimal investment and consumption under a continuous-time cointegration model with exponential utility. Quantitative Finance *19*, 1135–1149.

Mormann, F., Kreuz, T., Andrzejak, R.G., David, P., Lehnertz, K., and Elger, C.E. (2003). Epileptic seizures are preceded by a decrease in synchronization. Epilepsy Res. *53*, 173–185.

Oreskes, N. (2004). The scientific consensus on climate change. Science *306*, 1686.

Pedroni, P. (2001). Purchasing power parity tests in cointegrated panels. Rev. Econ. Stat. *83*, 727–731.

Pikovsky, A., Rosenblum, M., and Kurths, J. (2003). Synchronization: A Universal Concept in Nonlinear Sciences, *volume 12* (Cambridge university press).

Quiroga, R.Q., Kraskov, A., Kreuz, T., and Grassberger, P. (2002). Performance of different synchronization measures in real data: a case study on electroencephalographic signals. Phys. Rev. E *65*, 041903.

Robinson, P.M., and Hualde, J. (2003). Cointegration in fractional systems with unknown integration orders. Econometrica *71*, 1727–1766.

Shaw, J. (1981). An introduction to the coherence function and its use in EEG signal analysis. J. Med. Eng. Technol. *5*, 279–288.

Traub, R.D., and Wong, R. (1982). Cellular mechanism of neuronal synchronization in epilepsy. Science *216*, 745–747.

Tsay, R.S. (2005). Analysis of Financial Time Series, *volume 543* (John Wiley & Sons).

Tu, C. (2014). Cointegration-based financial networks study in Chinese stock market. Phys. Stat. Mech. Appl. *402*, 245–254.

Tu, C., DOdorico, P., and Suweis, S. (2018). Critical slowing down associated with critical transition and risk of collapse in cryptocurrency. arXiv, preprint. arXiv:1806.08386.

Turasie, A.A. (2012). Cointegration Modelling of Climatic Time Series, https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.574240.

Urquhart, A. (2017). Price clustering in bitcoin. Econ. Lett. *159*, 145–148.

Varela, F., Lachaux, J.P., Rodriguez, E., and Martinerie, J. (2001). The brainweb: phase synchronization and large-scale integration. Nat. Rev. Neurosci. *2*, 229.

Yang, C., Chen, Y., Niu, L., and Li, Q. (2014). Cointegration analysis and influence rank—a network approach to global stock markets. Phys. Stat. Mech. Appl. *400*, 168–185.

Zhang, X., Zwiers, F.W., Hegerl, G.C., Lambert, F.H., Gillett, N.P., Solomon, S., Stott, P.A., and Nozawa, T. (2007). Detection of human influence on twentieth-century precipitation trends. Nature *448*, 461.

ISCI, Volume *19*

Supplemental Information

Universal Cointegration and Its Applications

Chengyi Tu, Ying Fan, and Jianing Fan

Supplemental Figures

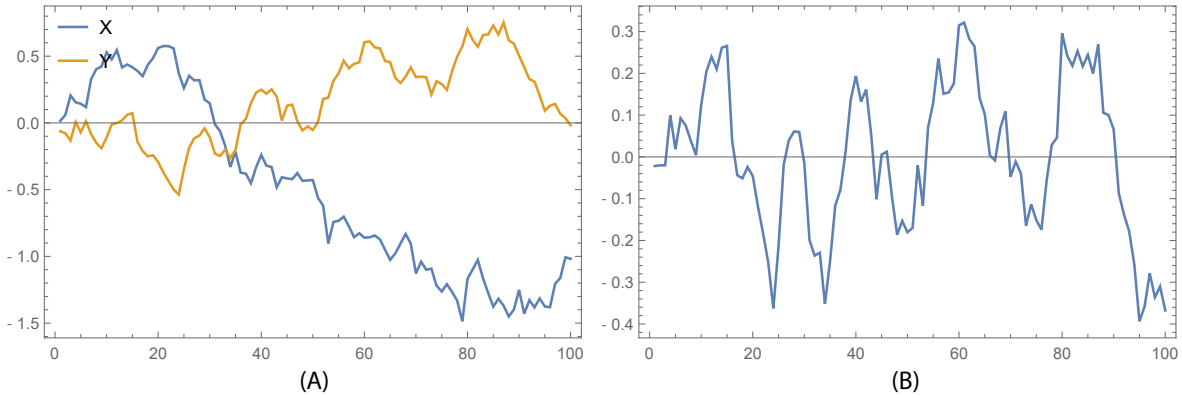


Figure S 1: Example of spurious regression, Related to Figure 1 and 2. (A) The time series of $x[k], y[k]$ generated by a random walk with random initial value. (B) The residual of the regression model.

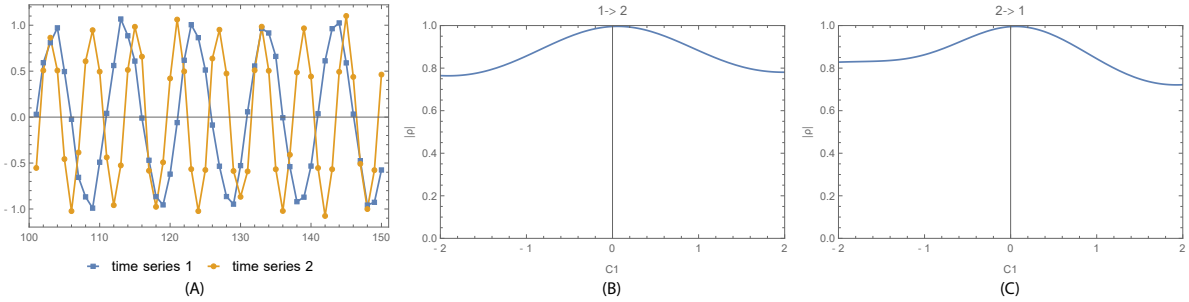


Figure S 2: Example to highlight the misuse of the classic cointegration test, Related to Figure 1 and 2. (A) The time series. (B-C) The acquisition function of Bayesian minimization for the time series shown in (A).

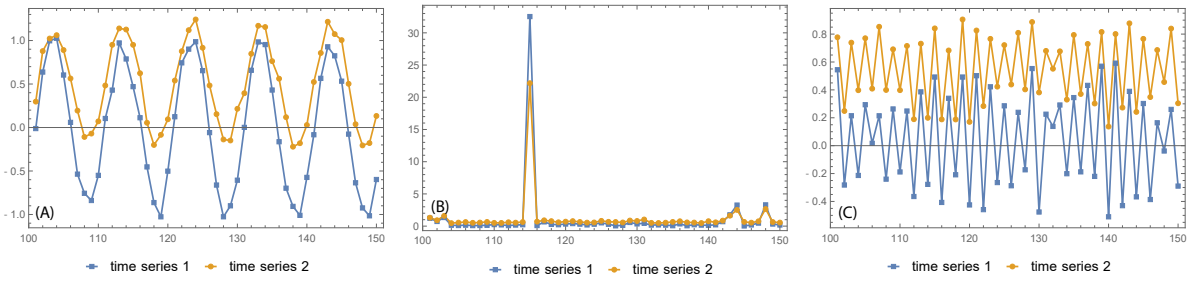


Figure S 3: Case of bidirectional coupling time series, Related to Figure 1. (A) Period oscillation; (B) extreme value; (C) chaotic map. The initial value is drawn from a uniform distribution between 0 and 1. Steps from 101 to 150 are adopted.

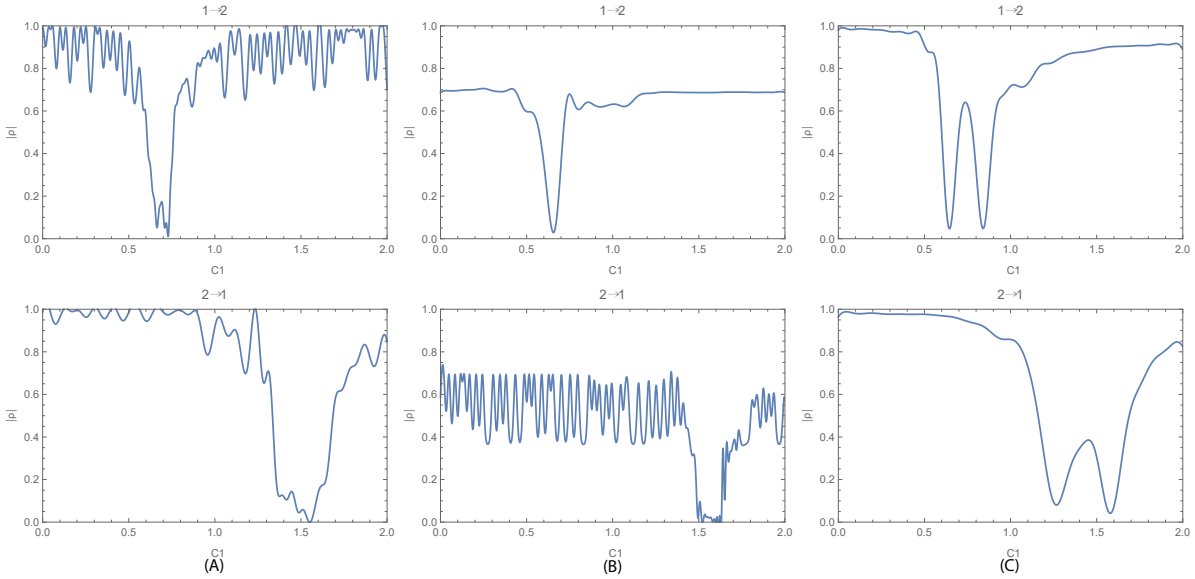


Figure S 4: The acquisition functions of Bayesian minimization for the three cases of bidirectional coupling time series shown in Figure S3, Related to Figure 2. (A) Period oscillation; (B) extreme value and (C) chaotic map. The first row checks the cointegrating relationship from times series 1 to 2, and the second row checks the opposite direction.

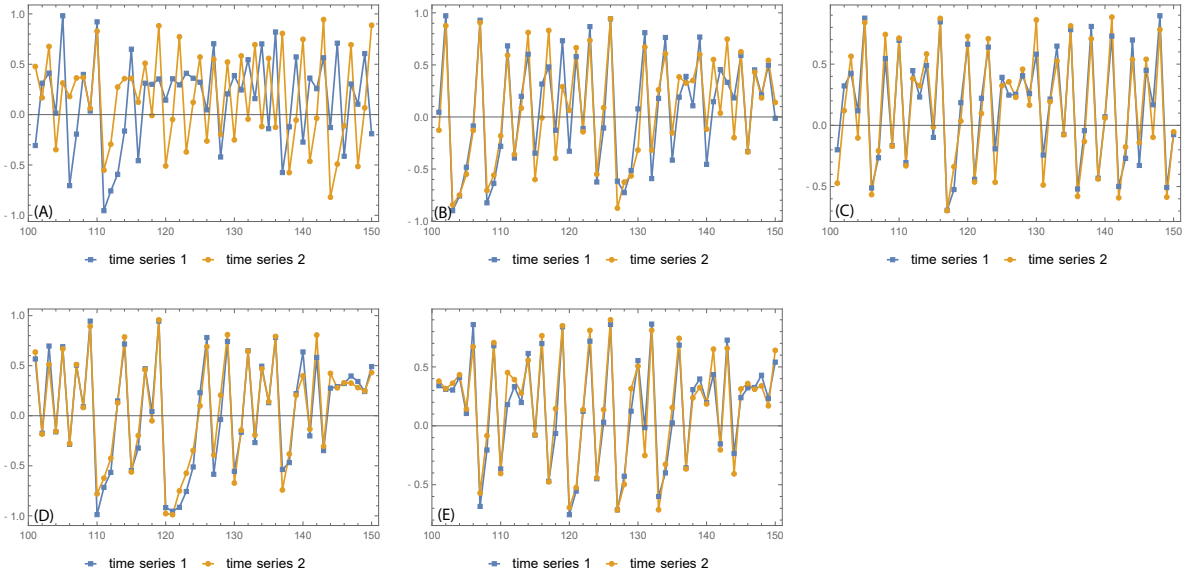


Figure S 5: Case of synchronization coupling time series with different coupling parameters, Related to Figure 1. (A) Coupling parameter 0.1; (B) coupling parameter 0.2; (C) coupling parameter 0.3; (D) coupling parameter 0.4; (E) coupling parameter 0.5.

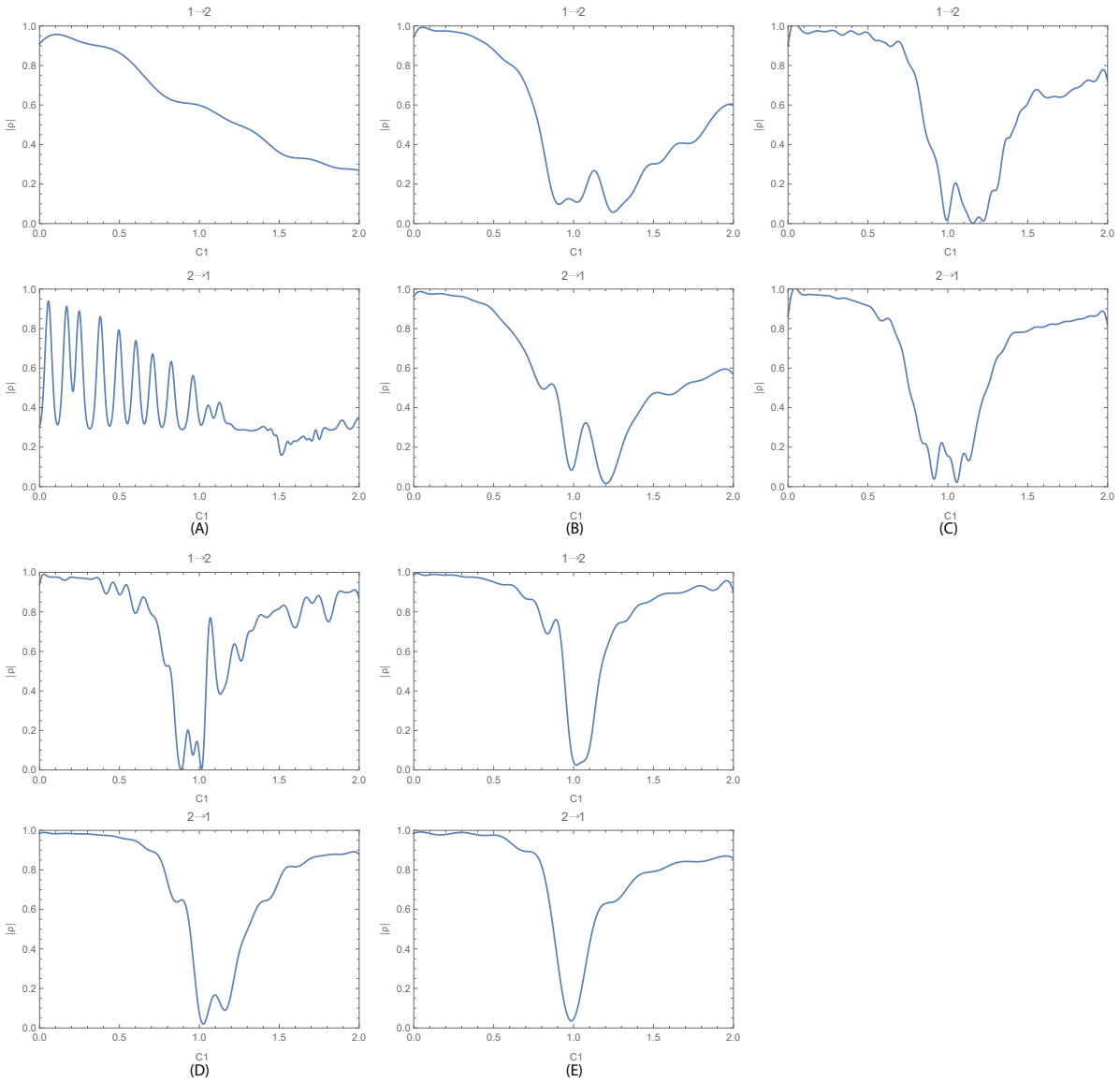


Figure S 6: The acquisition functions of Bayesian minimization for the synchronization coupling time series shown in Figure S5, Related to Figure 2. (A) Coupling parameter 0.1; (B) coupling parameter 0.2; (C) coupling parameter 0.3; (D) coupling parameter 0.4; (E) coupling parameter 0.5. For each case, we check the cointegrating relationship from times series 1 to 2 and vice versa.

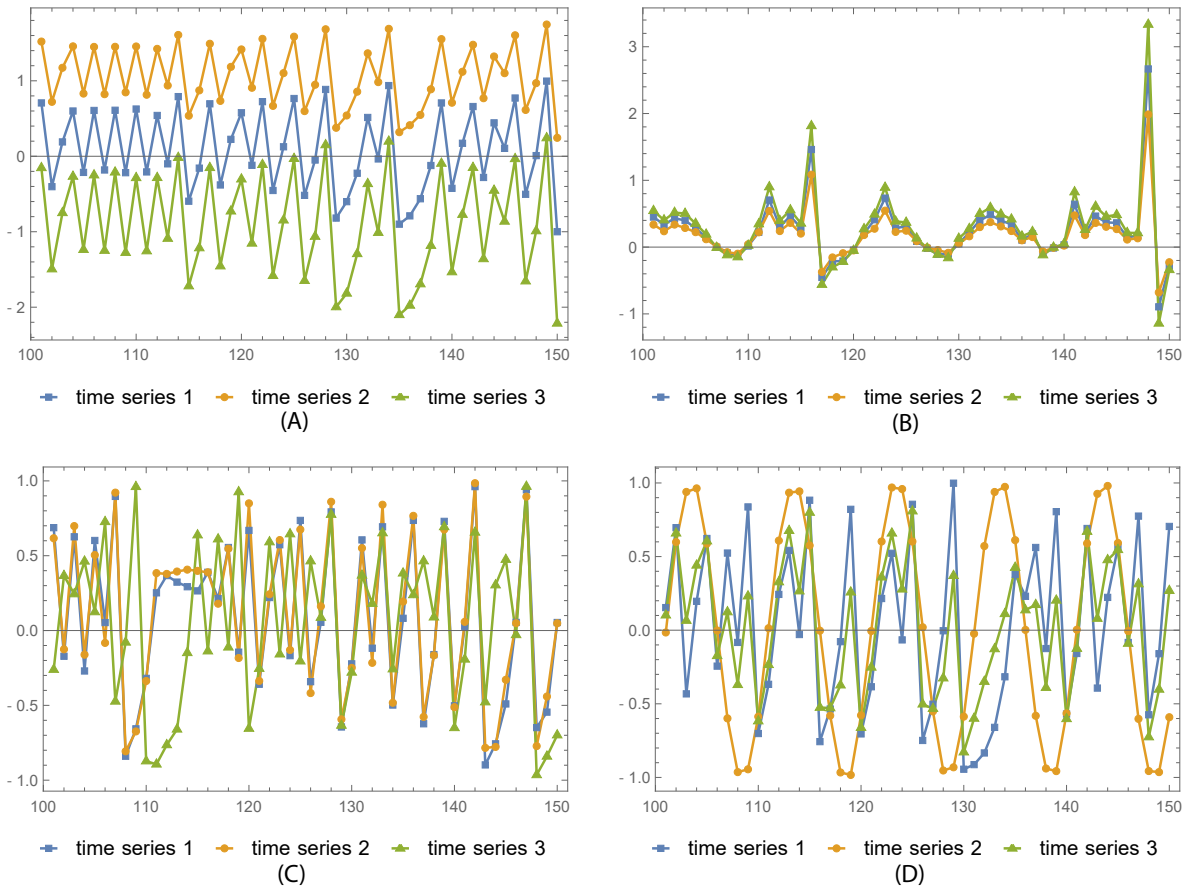


Figure S 7: Case of system with size larger than two, Related to Figure 1. (A) Unidirectional coupling; (B) bidirectional coupling; (C) synchronization coupling; (D) mixing coupling. The initial value is drawn from a uniform distribution between 0 and 1. Steps from 101 to 150 are adopted.

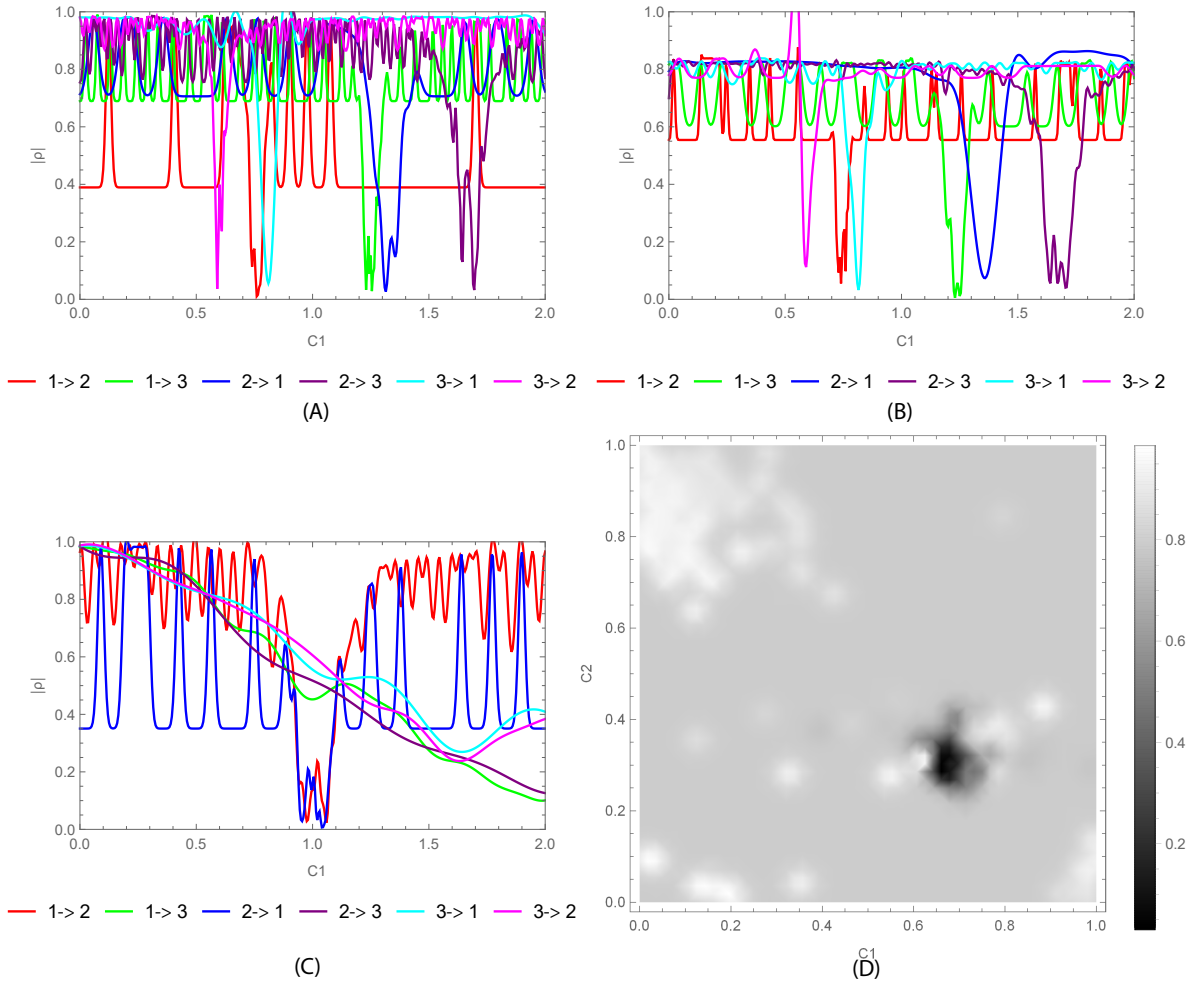


Figure S 8: The acquisition functions of the Bayesian minimization for coupling time series with number larger than 2 shown in Figure S7, Related to Figure 2. (A) Unidirectional coupling; (B) bidirectional coupling; (C) synchronization coupling; (D) mixing coupling. The density plot for mixing coupling where the gray level represents $|\rho|$, the absolute correlation of universal cointegration.

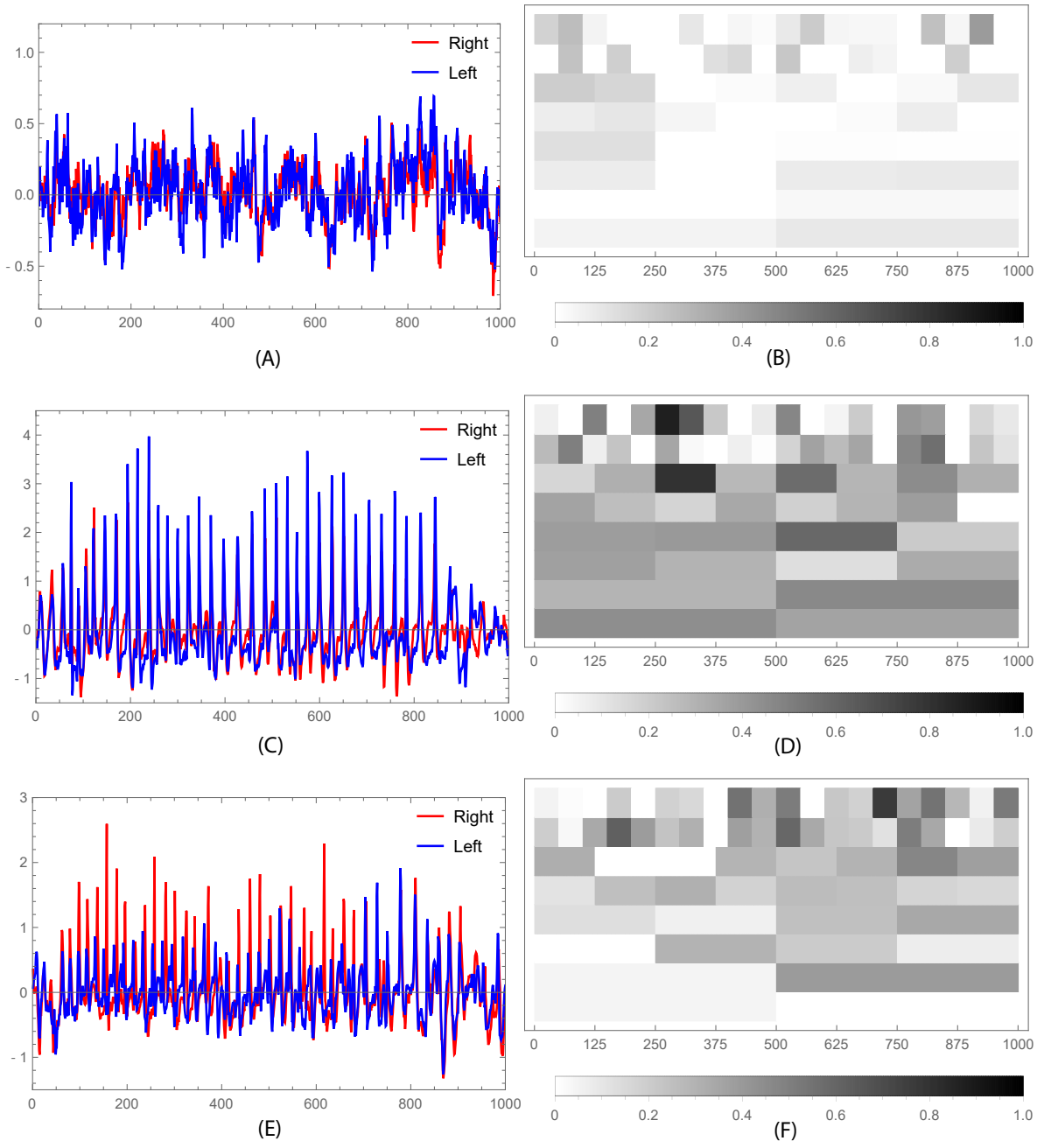


Figure S 9: The time series and absolute correlation of universal cointegration for electroencephalographic signals, Related to Figure 5. (A) The time series of case 1 without offset. (B) The absolute correlation $|\rho|$ of case 1 for different cell sizes. (C) The time series of case 2 without offset. (D) The absolute correlation $|\rho|$ of case 2 for different cell sizes. (E) The time series of case 3 without offset. (F) The absolute correlation $|\rho|$ of case 3 for different cell sizes. Each two rows represent a possible cointegration relationship for cell sizes of 50, 125, 250 and 500 from left to right and from right to left.

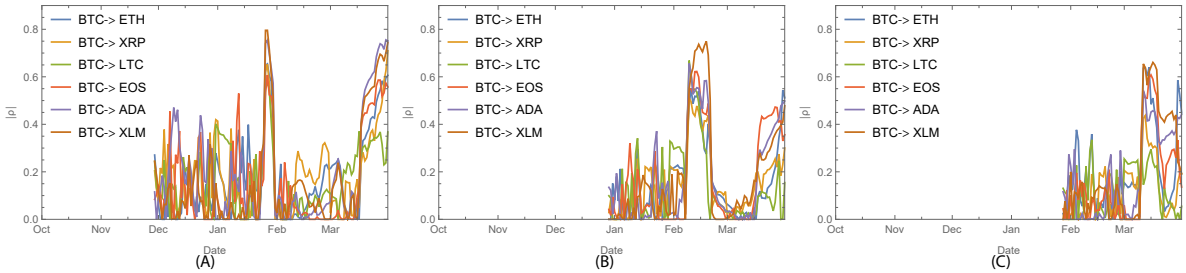


Figure S 10: The absolute correlation of the cointegration from BTC to other cryptocurrencies, Related to Figure 6. The time window is (A) 60, (B) 90 and (C) 120.

Supplemental Tables

Table S 1: The fitted parameter information of Figure S1, Related to Figure 1.

	Estimate	Standard Error	t-Statistic	p-Value
α	-0.0348	0.022851	-1.52282	0.131024
C1	-0.37769	0.027513	-13.7276	$1.47521 * 10^{-24}$

Table S 2: The stationary test of time series shown in Figure S1, Related to Figure 1. For each cell, the left is the p-value of the ADF test, and the right is the p-value of the KPSS test. The p-value of the ADF test is larger than the significance level of 0.05, and the p-value of the KPSS test is smaller than the significance level of 0.05; thus, it is $I(1)$.

	spurious regression case
time series 1	0.69411/0.01
time series 2	0.238397/0.01

Table S 3: The classic cointegration test of time series shown in Figure S1, Related to Figure 1. The second and third rows present the results of the Engle-Granger cointegration test. For each cell, the left is the p-value of the ADF test, and the right is the p-value of the PP test. Because the values are larger than the significance level of 0.05, cointegration relationships do not exist. The cells in the fourth and fifth rows present the results of the Johansen cointegration test. For each cell, the left is the p-value of the trace test, and the right is the p-value of the max eigenvalue test. Because both p-values are larger than the significance level of 0.05, the cointegration relationships do not exist.

	spurious regression case
$2 \rightarrow 1$	0.654737/0.654737
$1 \rightarrow 2$	0.486367/0.486367
$r \leq 0$	0.813227/0.82261
$r \leq 1$	0.586939/0.586939

Table S 4: The stationary test of time series shown in Figure 1 in the main text, Related to Figure 1. For each cell, the left is the p-value of the ADF test, and the right is the p-value of the KPSS test. The p-value of the ADF test is smaller than the significance level of 0.05, and the p-value of the KPSS test is larger than the significance level of 0.05; thus, both time series are stationary.

	period oscillation	extreme value	chaotic map
time series 1	$3.00 * 10^{-2}/1.00 * 10^{-1}$	$2.54 * 10^{-7}/1.00 * 10^{-1}$	$0.67 * 10^{-8}/1.00 * 10^{-1}$
time series 2	$1.53 * 10^{-1}/1.00 * 10^{-1}$	$1.51 * 10^{-7}/9.02 * 10^{-2}$	$1.12 * 10^{-5}/1.00 * 10^{-1}$

Table S 5: The classic cointegration test of time series shown in Figure 1 in the main text, Related to Figure 1. The second and third rows present the results of the Engle-Granger cointegration test. For each cell, the left is the p-value of the ADF test, and the right is the p-value of the PP test. Because the values are less than the significance level of 0.05, the cointegration relationships always exist. The cells in the fourth and fifth rows present the results of the Johansen cointegration. For each cell, the left is the p-value of the trace test, and the right is the p-value of the max eigenvalue test. Because the p-values are less than the significance level of 0.05, there are two cointegrating relationships.

	period oscillation	extreme value	chaotic map
$2 \to 1$	0.001/0.001	0.001/0.001	0.001/0.001
$1 \to 2$	0.001/0.001	0.001/0.001	0.001/0.001
$r \leq 0$	0.001/0.001	0.001/0.001	0.001/0.001
$r \leq 1$	0.034/0.034	0.001/0.001	0.001/0.001

Table S 6: The unit root test of the time series shown in Figure S2 (A), Related to Figure 1. For each cell, the left is the p-value of the ADF test, and the right is the p-value of the KPSS test. The p-value of the ADF test is smaller than the significance level of 0.05, and the p-value of the KPSS test is larger than the significance level of 0f.05; thus, it is $I(0)$.

	misuse case
time series 1	0.030641/0.1
time series 2	0.000109/0.1

Table S 7: The classic cointegration test of the time series shown in Figure S2 (A), Related to Figure 1. The second and third rows present the results of the Engle-Granger cointegration test. For each cell, the left is the p-value of the ADF test, and the right is the p-value of the PP test. Because the values of the first row are larger than the significance level of 0.05 and the second row are smaller, the cointegrating relationship from time series 2 to 1 does not exist, but 1 to 2 exists. The cells in the fourth and fifth rows present the results of the Johansen cointegration test. For each cell, the left is the p-value of the trace test, and the right is the p-value of the max eigenvalue test. Because all p-values are smaller than the significance level of 0.05, the cointegration relationship does not exist.

	misuse case
$2 \to 1$	0.463512/0.463512
$1 \to 2$	0.009955/0.009955
$r \leq 0$	0.012055/0.03913
$r \leq 1$	0.032891/0.032891

Table S 8: The minimum and its argument for the unidirectional coupling shown in Figure 2 of the main text, Related to Figure 2.

	period oscillation	extreme value	chaotic map
$1 \to 2$	0.0372/-1.46	0.00239/-1.45	0.0395/-1.48
$2 \to 1$	0.0130/-0.688	0.000940/-0.613	0.00803/-0.702

Table S 9: The minimum and its argument for bidirectional coupling time series shown in Figure S3, Related to Figure 2. For each cell, the left is the minimum, and the right is its argument.

	period oscillation	extreme value	chaotic map
$1 \to 2$	0.0357/0.732	0.0613/0.669	0.0497/0.649
$2 \to 1$	0.00910/1.56	0.000912/1.60	0.0416/1.58

Table S 10: The minimum and its argument for the synchronization coupling time series shown in Figure S5, Related to Figure 2. For each cell, the left is the minimum, and the right is its argument.

	0.1	0.2	0.3	0.4	0.5
$1 \to 2$	0.287/1.83	0.0996/0.900	0.00231/1.16	0.00620/1.01	0.0308/1.03
$2 \to 1$	0.176/1.50	0.0845/0.983	0.0219/1.06	0.0598/1.05	0.0743/1.02

Table S 11: The $|\rho|$, C1, α and distance $\sqrt{(\rho)^2 + (C_1 - 1)^2 + (\alpha)^2}$ for all 48 models, Related to Figure 4.

model	$\|\rho\|$	C1	α	distance
ACCESS1-0	0.012342	1.29406	-0.1506	0.330613
ACCESS1-3	0.268489	1.49518	-0.13692	0.579683
bcc-csm1-1	0.070241	0.669578	-0.10032	0.352386
bcc-csm1-1-m	0.152546	0.829021	-0.14488	0.271097
BNU-ESM	0.058741	0.725356	-0.16815	0.327345
CanESM2	0.093166	1.10394	-0.17373	0.222859
CCSM4	0.016328	0.779267	-0.16853	0.278192
CESM1-BGC	0.157871	0.899118	-0.16473	0.249469
CESM1-CAM5	0.035197	1.49518	-0.169	0.524404
CESM1-CAM5-1-FV2	0.046752	1.21319	-0.0846	0.234081
CESM1-FASTCHEM	0.068458	0.850536	-0.192	0.252765
CESM1-WACCM	0.020253	0.811687	-0.16184	0.249125
CMCC-CM	0.089119	1.32915	-0.13865	0.368114
CMCC-CMS	0.021925	1.22725	-0.08051	0.242088
CMCC-CESM	0.160444	1.35373	-0.07218	0.395064
CNRM-CM5	0.095295	1.14276	-0.14872	0.227108
CSIRO-Mk3-6-0	0.054235	1.41373	-0.2031	0.464076
EC-EARTH	0.027116	1.12038	-0.18928	0.225952
FGOALS-g2	0.019714	1.14144	-0.12335	0.188706
FIO-ESM	0.084527	1.00668	-0.17448	0.193993
GFDL-CM3	0.032854	1.33445	-0.0996	0.350513
GFDL-ESM2G	0.045564	0.940014	-0.12067	0.142255
GFDL-ESM2M	0.120086	1.40847	-0.17551	0.460513
GISS-E2-H_p1	0.021468	1.16451	-0.25464	0.303917
GISS-E2-H_p2	0.219712	1.3511	-0.23713	0.477259
GISS-E2-H_p3	0.042612	1.06338	-0.2295	0.241875
GISS-E2-H-CC_p1	0.040999	0.799873	-0.17	0.265765
GISS-E2-R_p1	0.017551	1.17746	-0.16638	0.24389
GISS-E2-R_p2	0.01071	1.30174	-0.18495	0.354077
GISS-E2-R_p3	0.13759	1.2539	-0.18594	0.343469
GISS-E2-R-CC_p1	0.199379	1.00857	-0.14044	0.244026
HadGEM2-AO	0.039472	1.13624	-0.1305	0.192739
HadGEM2-CC	0.018416	1.4892	-0.03746	0.49098
HadGEM2-ES	0.016071	1.42992	-0.10598	0.443083
inmcm4	0.00113	1.1967	-0.16774	0.258511
IPSL-CM5A-LR	0.074578	0.798226	-0.18203	0.281798
IPSL-CM5A-MR	0.002509	1.04942	-0.20224	0.20821
IPSL-CM5B-LR	0.019022	1.21838	-0.26961	0.347474
MIROC5	0.035507	1.45188	-0.12077	0.469087
MIROC-ESM	0.00451	1.17831	-0.19744	0.266076
MIROC-ESM-CHEM	0.117915	1.19558	-0.18613	0.294616
MPI-ESM-LR	0.115608	1.03706	-0.18033	0.217392
MPI-ESM-MR	0.132282	0.946505	-0.17015	0.222064
MPI-ESM-P	0.194011	1.05722	-0.23278	0.308386
MRI-CGCM3	0.323931	1.4992	-0.17633	0.620661
MRI-ESM1	0.106964	1.49988	-0.19227	0.546162
NorESM1-M	0.056273	1.28401	-0.14899	0.325615
NorESM1-ME	0.142396	1.23904	-0.14835	0.315317

Transparent Methods

1. Spurious regression

Although applying regression models to time series is heuristic thinking, there are many fatal disadvantages. The most well-known disadvantage is that spurious regression may mislead hypothesis testing (Granger et al., 2001). Spurious regression is a regression that provides misleading statistical evidence of a linear relationship between independent nonstationary variables. To better illustrate the phenomenon of spurious regression, we first generate two different random walk time series $x[k+1] = x[k] + \epsilon_x, y[k+1] = y[k] + \epsilon_y$, where ϵ_x, ϵ_y are two normal distributions with a mean of 0 and a standard deviation of 0.05 (see Figure S1 (A)). Then, we regress the time series $x\,[k]$ onto time series $y\,[k]$. Because the two time series are independent, (i) the coefficient should be close to zero, (ii) the t-value is most often insignificant, and (iii) R^2 is typically very low. However, these results do not appear, and the regression indicates a nonexisting relationship (see Figure S1 (B) and Tabure S1). In general, classic statistical methods, such as ordinary least squares (OLS), are based on the assumption that the time series are stationary processes. Therefore, if time series are nonstationary, it is incorrect to use OLS to estimate their long-term linear relationships.

2. Background of cointegration

Cointegration focuses on whether the long-term linear relationship between two or more time series is stationary even if this linear relationship does not exist or is not strong for the short term. A time series is said to be (weakly) stationary if its mean and autocovariance do not vary with respect to time, i.e., (i) $E\,[x[k]] = \mu < \infty$ for all $k \in N$ and (ii) $E\,[(x[k]-\mu)(x[k+h]-\mu)] = C(h)$ for all $k \in N$ and $h \in N_0$. The existence of a cointegrating relationship between time series implies that they have a common stochastic drift in the long term.

The classic cointegration only considers the case $CI(d,b)$ (Hamilton, 1994). For a nonstationary time series, the minimum number of differences required to obtain a stationary series is called the order of integration, denoted as $I(d)$, where d is the minimum number of differences. The vector time series $\mathbf{x}[k]$ are said to be cointegrated of order $CI(d,b)$, where $d > b > 0$, if (i) all time series of vector $\mathbf{x}[k]$ are $I(d)$, and (ii) a vector $\mathbf{C} \neq 0$ exists such that $\mathbf{C}^T\mathbf{x}[k]$ is $I(d-b)$. The vector $\mathbf{C}$ is called the cointegrating vector, and $\mathbf{C}^T\mathbf{x}[k]$ is called the cointegrating relationship. In particular, financial and economic time series are always assumed to be $I(1)$, also called unit root. In most cases, their linear combination is still $I(1)$. However, in some cases, the linear combination becomes a $I(0)$ time series. Classic cointegration always identifies and searches for this relationship. The main methods are the Engle-Granger cointegration test and Johansen cointegration test. However, there are many fatal drawbacks in the classic cointegration analysis:

(a) The vector $\mathbf{x}[k]$ is assumed to be composed of time series with the same order of integration. In economics, people always assume that each time series is $I(1)$. Otherwise, they will adopt preprocessing, such as difference or log-transform. However, these preprocessings cannot guarantee that all processed time series are all $I(1)$. For example, the orders of integration of some time series are still different after preprocessing. Regardless, the following cointegration test is applied to the processed time series rather than the original time series. Therefore, the result is not reliable and clear.

(b) It is difficult or even impossible to address time series where any preprocessing cannot obtain the $I(1)$: (i) If a time series is a cycle, for example, sin function, limited difference cannot be guaranteed to obtain the $I(1)$ time series (see Figure 1 (A) in the main text); (ii) If a time series includes some peaks composed of extreme values, most preprocessings fail (see Figure 1 (B) in the main text); (iii) If a time series is generated by chaos, it is stationary (see Figure 1 (C) in the main text).

(c) It is invalid for complete synchronization coupling composed of time series from a chaotic system. If the time series is generated by a completely synchronized chaotic system, any stationary test will report that it is stationary, and any preprocessing cannot obtain the corresponding $I(1)$ time series. The values of each step are the same or at least similar, but their relationship does not exactly exist.

Consequently, it is important in theory and practice to introduce a method that can (a) be applied to time series whose order of integration is not 1, (b) identify the possible cointegration in the original time series rather than the processed time series, and (c) detect cases of the system undergoing complete synchronization coupling composed of time series from a chaotic system.

3. Classic cointegration analysis

The classic cointegration analysis mainly includes the Engle-Granger cointegration test (Engle and Granger, 1987) and Johansen cointegration test (Johansen, 1995), after Engle and Granger realized the misuse of the regression model to check long-term relationship (Engle and Granger, 1987) and proposed the concept of cointegration. These methods only consider the cointegration $CI(1,1)$ (Tsay, 2005), i.e., (i) all time series of vector $\mathbf{x}[k]$ are $I(1)$, also called unit root, and (ii) a vector $\mathbf{C} \neq 0$ exists such that $\mathbf{C}^T\mathbf{x}[k]$ is $I(0)$. The prerequisite of these methods is that each given time series should be $I(1)$, namely, it is nonstationary, and the minimum number of differences required to obtain a stationary series is 1. Here, we report the whole process of classic cointegration analysis.

3.1. $I(1)$ test

Before applying the classic cointegration test, the time series should be guaranteed to be $I(1)$. There are many tests for this goal. The common criterion is mixing the results of the augmented Dickey-Fuller (ADF) test (Said and Dickey, 1984) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test (Kwiatkowski et al., 1992). The null hypothesis of the ADF test is that the given time series is unit root. If its p-value is larger than the significance level, the null hypothesis is accepted, i.e., its order of integration is 1. Unlike the ADF test, the KPSS test is actually a stationarity test, i.e., its null hypothesis is that the given time series is stationary. If its p-value is larger than the significance level, the null hypothesis is accepted, i.e., its order of integration is 0. For the given time series, if its ADF test is accepted and the KPSS test is rejected, its order of integration is and only is 1.

3.2. Engle-Granger cointegration test

The Engle-Granger cointegration test, introduced by Engle and Granger, is the first and most well-known cointegration test. It searches the linear combination having minimum variance based on ordinary least squares (Engle and Granger, 1987). Assuming that one-dimensional time series x_t and one-dimensional time series y_t are $I(1)$, the Engle-Granger cointegration test attempts to check whether they are $CI(1,1)$. The procedure is as follows: (i) estimating a regression by OLS, i.e., a parameter β exists such that $y_t = \beta x_t + u_t$; (ii) testing the residual series μ_t for weak stationarity by some tests, such as the ADF test (Said and Dickey, 1984) and Phillips-Perron (PP) test (Phillips and Perron, 1988). If $y_t - \beta x_t = u_t \sim I(0)$ is a well-behaved equilibrium, the existence of cointegration is manifested; otherwise, the equilibrium does not exist. For the ADF or PP test, the null hypothesis is that μ_t is still $I(1)$, i.e., cointegration does not exist. If its p-value is larger than the significance level, the null hypothesis is accepted, and then the cointegrating relationship does not exist. Additionally, if the time series $\mathbf{x}_t$ is higher dimensional, the procedure is similar, i.e., searching the parameter vector β such that $y_t = \beta\mathbf{x}_t + u_t$.

3.3. Johansen cointegration test

The standard approach for cointegration test is the Johansen cointegration test. This test searches for the linear combination that is the most stationary based on the maximum likelihood estimator of the so-called reduced rank model (Johansen, 1995). Considering a vector autoregression model (VAR) of order k, $x_t = \Pi_1 x_{t-1} + \cdots + \Pi_{k-1}x_{t-k+1} + \Pi x_{t-k} + \mu + \epsilon_t$, where x_t is $I(1)$ time series, $\Pi_i, i = 1, \ldots, k$ describes short-term changes resulting from previous changes, μ is a constant vector, and ϵ_t is a white noise. Its vector error correction (VEC) model is $\Delta x_t = \Gamma_1\Delta x_{t-1} + \cdots + \Gamma_{k-1}\Delta x_{t-k+1} + \Pi x_{t-k} + \mu + \epsilon_t$, where $\Delta x_t = x_t - x_{t-1}$, $\Gamma_1 = \Pi_1 - I$, $\Gamma_2 = \Pi_2 - \Gamma_1, \ldots, \Gamma_k = \Pi_k - \Gamma_{k-1}$ and $\Pi = I - \Pi_1 - \Pi_2 - \cdots - \Pi_k = \alpha\beta\prime$, where α is a $p \times k$ matrix to describe the changes that help restore an equilibrium and β is a $p \times k$ matrix to characterize the long-term relationships between levels of variables. Because Δx_t is stationary, Πx_{t-k} must be stationary for a meaningful equation. Let $Z_{0t} = \Delta x_t$, $Z_{1t} = (\Delta x\prime_{t-1}, \ldots, \Delta x\prime_{t-k+1}, 1)\prime$ and $Z_{kt} = x_{t-k}$; their product moment matrix is calculated as $M_{ij} = T^{-1}\sum_{t=1}^{T} Z_{it}Z\prime_{jt}(i, j = 0, 1, k)$. By denoting the residual sum of squares from regressing Z_0 and Z_k on Z_1 as S_{ij}, where $i, j = 0, k$, the error terms are calculated as $S_{ij} = M_{ij} - M_{i1}M_{11}^{-1}M_{1j} = \frac{1}{T}\sum_{t=1}^{T} R_{it}R\prime_{jt}$. The equilibrium relation matrix is $\Pi = [S_{kk}^{-1}S_{k0}S_{00}^{-1}S_{0k}]$, and its rank provides the number of different cointegration vectors between variables and the cointegration degree. After Π matrix is formed, nonzero $\hat{\lambda}_i$ eigenvalues should be calculated in descending order. If $r = k$, then x_t is stationary. If $r = 0$, then Δx_t is stationary, and it is not possible to obtain stationary relations between the levels of the variables by linear combinations. If $k > r > 0$, then $\Delta x_t \sim I(1)$, and r cointegrating vectors exist in which the process can be made stationary by linear combinations, $\beta\prime x_t$. There are two likelihood ratios to test the relationship between the rank r and order k. The trace test, $J_{trace} = -T\sum_{i=r+1}^{n} ln(1 - \hat{\lambda}_i)$, where T is the sample size and $\hat{\lambda}_i$ is the i-th largest canonical correlation, tests the null hypothesis of r cointegrating vectors against the

alternative hypothesis of n cointegrating vectors. The maximum eigenvalue test, $J_{max} = -T(1 - \hat{\lambda_{r+1}})$, tests the null hypothesis of r cointegrating vectors against the alternative hypothesis of $r + 1$ cointegrating vectors. Finally, r different cointegrating vectors can be obtained. Note that the order of integration of each variable should be 1, i.e., $I(1)$ before applying the Johansen cointegration test. For the trace and maximum eigenvalue test, the null hypothesis is a cointegrating rank less than or equal to r. If the p-value is larger than the significance level, the null hypothesis is accepted, i.e., the number of cointegrating relationships is less than or equal to r.

4. Application of classic cointegration and its misuse

We emphasize that even if the prerequisite of the Engle-Granger cointegration test and Johansen cointegration test is not satisfied, i.e., all time series are all $I(1)$, they can still be calculated mechanically. Although they occasionally can obtain the real cointegrating relationship in the universal sense, the methods are not reliable, and there is no necessary connection between these test results and cointegration as defined by Engle and Granger. We first apply the classic cointegration to overcome the spurious regression. Then, we apply it to some cases that fall outside of its range. Finally, we show an example to disclose its dangers and call on avoiding its misuse.

4.1. The classic cointegration analysis to overcome the spurious regression

If time series are nonstationary, adopting OLS to estimate the long-term linear relationships will cause spurious regression (see Table S1). Furthermore, if they are $I(1)$ (see Table S2), the classic cointegration can overcome the spurious regression (see Table S3).

4.2. Example of the classic cointegration analysis

Here, we apply classic cointegration analysis to the first model example in the main text. Although these examples fall outside of the range of classic cointegration (see Table S4), they still determine the cointegration in the universal sense (see Table S5).

4.3. The misuse of classic cointegration test

Here, we show an example to highlight the dangers of classic cointegration analysis and call on avoiding its misuse. Again, this analysis is not a criticism of the classic cointegration, as this example falls outside of its range. We note that it occasionally may be possible to find a cointegration relationship. However, to do so, one must know in advance that the assumption of the classic cointegration is not satisfied. Thus, even if the results do not invalidate these tests, the ambiguity of multiple results would make them useful only post hoc and not as a way to determine cointegration a priori when the answer is not already known.

Considering the general model, where A is the identity matrix with size 2, i.e., $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, coupling dynamics $G_1(x_1[k]) = \mathrm{Sin}(\frac{\pi k}{5})$ and $G_1(x_1[k]) = \mathrm{Cos}(\frac{\pi k}{3})$, the process noise and observation noise of each component are both $N(0, 0.05^2)$, the generated time series are independent (see Figure S2 (A)) and $I(0)$ (see Table S6). The Engle-Granger cointegration test rejects the cointegrating relationship from time series 2 to 1 but accepts the cointegrating relationship from 1 to 2; the Johansen cointegration test rejects all possible cointegrating relationships although its p-value approximates to the significance level (see Table S 7). The result is contradictory and ambiguous. According to universal cointegration, the absolute correlation is always high (see Figure S2 (B, C)); thus, the cointegration clearly does not exist.

5. Framework of universal cointegration

Our universal cointegration returns to the essence of the cointegration definition, only considering whether the long-term linear relationship between two or more time series is stationary and its extent. In Eq. 4 of main text, if $y_1[k], y_2[k]$ are $I(1)$, the classic cointegration can be applied (see Transparent Methods, section Classic cointegration analysis); otherwise, it fails. Irrespective of the integration or even the type of the time series, universal cointegration can indicate whether the long-term linear relationship is stationary even if short-term deviations exist, its extent and the corresponding cointegrating vector.

Considering a vector $\mathbf{x}[k]$ composed of time series $x_N[k]$ as independent and $x_i[k],\ i \in 1, \ldots, N-1$ as dependent; if and only if the equation $x_N[k] = \alpha + C_1 x_1[k] + \cdots + C_{N-1} x_{N-1}[k] + \epsilon[k]$ holds, where $\alpha, C_1, \ldots, C_{N-1}$ are

constant and $\epsilon[k]$ is a random noise, is the long-term linear relationship stationary, and $x_N[k] - C_1x_1[k] - \cdots - C_{N-1}x_{N-1}[k] = \epsilon[k]$ is a random noise with mean α. This relationship is called universal cointegration. We adopt convergent cross mapping (CCM) (Sugihara et al., 2012; McCracken and Weigel, 2014; Tsonis et al., 2018) to determine whether $x_N[k] - C_1x_1[k] - \cdots - C_{N-1}x_{N-1}[k]$ is a random noise (see Transparent Methods, section Convergent Cross Mapping). If and only if time series X is causally influencing time series Y does Y include the information of X. CCM tests for causation by measuring the extent to which the state of Y can reliably estimate the states of X. The more the correlation of CCM approximates to 1, the stronger is the causality, i.e., Y includes more information of X. By contrast, if the correlation of CCM approximates to 0, then there is no causality from X to Y, i.e., Y does not include any information of X. If vector $\mathbf{C} = \{C_1, \ldots, C_{N-1}, -1\} \neq 0$ exists such that $|\rho|$, the absolute correlation of CCM between cause $x_N[k]$ and result $\mathbf{C}^T\mathbf{x}[k]$, approximates to 0, then $x_N[k] - C_1x_1[k] - \cdots - C_{N-1}x_{N-1}[k]$ will not include any information of $x_N[k]$. Furthermore, if $x_N[k]$ includes information of other unknown time series, $|\rho|$ cannot approximate to 0 regardless of the vector $\mathbf{C}$. Therefore, when $|\rho| \to 0$, $x_N[k] - C_1x_1[k] - \cdots - C_{N-1}x_{N-1}[k]$ does not include any information, i.e., it is a random noise, the universal cointegration exists. In brief, the universal cointegration and its extent map to searching the vector $\mathbf{C}$ to minimize the absolute correlation of CCM between cause $x_N[k]$ and result $\mathbf{C}^T\mathbf{x}[k]$. If the minimum approximates to 0, the universal cointegration exists, and the corresponding vector $\mathbf{C}$ is the desired cointegrating vector and α is the mean of all steps $x_N[k] - C_1x_1[k] - \cdots - C_{N-1}x_{N-1}[k]$ ($\alpha = \langle x_N[k] - C_1x_1[k] - \cdots - C_{N-1}x_{N-1}[k] \rangle$). The smaller $|\rho|$ is, the stronger the universal cointegration is.

We adopt Bayesian minimization (Brochu et al., 2010; Snoek et al., 2012) to search the desired vector corresponding to the minimum of the absolute correlation of CCM because of its natural merits (see Transparent Methods, section Bayesian minimization): (i) It can offer not only the minimum but also the acquisition function as a proxy of the expensive objection function. Therefore, we can determine the minimum according to the monotonicity of the acquisition function easily and robustly. (ii) Whether the possible universal cointegration is valid depends on how small of the absolute correlation is sufficient. The acquisition function of Bayesian minimization can offer all concave vertexes. (iii) When the given time series is more than 2, Bayesian minimization can provide us all possible cointegrating relationships and their extent. Considering three time series $x_1[k], x_2[k], x_3[k]$, if the cointegrating relationship between $x_1[k]$ and $x_2[k]$ exists, the density plot of the acquisition function of the Bayesian minimization, i.e., the absolute correlation between cause $x_3[k]$ and result $x_3[k] - C_1x_1[k] - C_2x_2[k]$ as a function of C_1, C_2, will be one line; otherwise, it will just be one point.

If the number of time series $\mathbf{x}[k]$ is too large, the calculation of Bayesian minimization is hard or even unrealized. We can apply CCM as a preprocessing to remove some time series without effective information. Considering the general case, $\mathbf{x}[k]$ composed of N time series $x_i[k]$, $i \in 1, \ldots, N$, if the absolute correlation of CCM between cause $x_N[k]$ and result $x_i[k]$ approximates to 0, the result $x_i[k]$ does not include information of $x_N[k]$, i.e., the possible cointegrating relationship does not need to include this time series $x_i[k]$. Therefore, $x_i[k]$ can be removed from the possible cointegrating relationship before performing the Bayesian minimization.

6. Convergent Cross Mapping

Recently, Sugihara et al. introduced a novel method, convergent cross mapping (CCM) (Sugihara et al., 2012; McCracken and Weigel, 2014; Tsonis et al., 2018), to identify the causal relationships from observational data alone (Sugihara et al., 2012; Pearl, 2009). If and only if time series X is causally influencing time series Y does Y include the information of X. CCM tests for causation by measuring the extent to which the state of Y values can reliably estimate the states of X. The more the correlation of CCM approximates to 1, the more information of X that is in Y.

According to Takens' theorem (Takens, 1981), in dynamic systems, if one time series $y\,[k]$ influences another time series $x\,[k]$, then $x\,[k]$ will include the information of $y\,[k]$. In other words, if $y\,[k]$ is indeed causal to $x\,[k]$, $x\,[k]$ can reconstruct $y\,[k]$. CCM measures the extent to which the time delay embedding of $x\,[k]$ can reliably estimate the state $y\,[k]$. In practice, the Pearson correlation coefficient between the original time series $y\,[k]$ and its estimate from the CCM by another time series $x\,[k]$ is adopted as the criterion. The larger the correlation, the better the $x\,[k]$ reconstructs $y\,[k]$, and then the stronger the causal effect.

For the time series $x\,[k]$ and $y\,[k]$, where $k = 1, \ldots, L$, the CCM algorithm to test the causality and its extent from $x\,[k]$ to $y\,[k]$ is shown as follows (McCracken and Weigel, 2014):

Step 1 For the given embedding dimension E, the lagged-coordinate vectors of $x\,[k]$ are

$\tilde{x}[k] = \{x[k], x[k-\tau], \ldots, x[k-(E-1)\tau]\}$, where the first and last are created at $k = 1+(E-1)\tau$ and $k = L$. The set of vectors $\tilde{x}[k]$ is its shadow manifold M_x;

Step 2 Because the minimum number of points required for a bounding simplex in an E-dimensional space is E + 1, for each vector $\tilde{x}[k]$, we search its E+1 nearest neighbors $k_1, k_2, \ldots, k_{E+1}$ in the shadow manifold M_x, and they are ordered by the Euclidean distance $d_i = D(\tilde{x}[k], \tilde{x}[k_i])$, where $i = 1, \ldots, E+1$;

Step 3 The weight of each nearest neighbor is calculated as $\omega_i = u_i / \sum_{j=1}^{E+1} u_j$, where $u_i = e^{-d_i/d_1}$;

Step 4 The estimate of each point $y[k]$ is calculated as $\hat{y}[k]|M_x = \sum_{i=1}^{E+1} \omega_i y[k_i]$;

Step 5 The CCM correlation is defined as the standard Pearson's correlation coefficient $\rho(y[k], \hat{y}[k]|M_x)$. If the causality is strong, $\hat{y}[k]|M_x$ should converge to $y[k]$, i.e., $\rho(y[k], \hat{y}|M_x) \to 1$.

7. Bayesian minimization

Many scientific papers have focused on the problem of optimizing a nonlinear function over a given set (Brochu et al., 2010; Shahriari et al., 2016). Consider a function $f(\mathbf{x})$ and that the problem is briefly expressed by $\mathbf{x}^* = \arg\min_{\mathbf{x}\in\mathcal{X}} f(\mathbf{x})$, where $\mathbf{x} \in \mathcal{X}$ is the realm of optimization. When evaluating the objective function, $f(\mathbf{x})$ is expensive, and the derivatives and convexity properties are unknown, such as machine-learning model (Snoek et al., 2012; Robert, 2014), drug trials (Haines et al., 2003) and neuroscience (Friston, 2012; Shi et al., 2013). Bayesian optimization is a powerful strategy for finding the extrema of objective functions. Bayesian optimization employs the Bayesian technique of setting a prior over the objective function and combining it with evidence to obtain a posterior function (Brochu et al., 2010; Snoek et al., 2012). This permits a utility-based selection of the next observation to make on the objective function, which must take into account both exploration (sampling from areas of high uncertainty) and exploitation (sampling areas likely to offer improvement over the current best observation).

Assuming that $\mathbf{x}_i$ is the i-th sample and $f(\mathbf{x}_i) = f_i$ is its observation of the objective function, Bayesian optimization adopts an appropriate Gaussian process (GP) prior to explain the objective function, i.e., $P(f) = \mathcal{GP}(f; \mu, K)$. For convenience, we assume that the prior mean is zero and that the covariance function is the squared exponential function $k(x_i, x_j) = \exp(-1/2||x_i - x_j||^2)$. Therefore, the function values are drawn according to a multivariate normal distribution $\mathcal{N}(\mathbf{0}, \mathbf{K})$, where the kernel matrix is given by $\mathbf{K} = \begin{bmatrix} k(x_1,x_1) & \ldots k(x_1,x_t) \\ \vdots & \vdots \\ k(x_t,x_1) & \ldots k(x_t,x_t) \end{bmatrix}$. If we already have the observations $\{\mathbf{x}_{1\ldots t}, \mathbf{f}_{1\ldots t}\}$ from previous iterations, where $\mathbf{x}_{1\ldots t} = \{x_1, \ldots, x_t\}$ and $\mathbf{f}_{1\ldots t} = \{f_1, \ldots, f_t\}$, $\mathbf{f}_{1\ldots t}$ and f_{t+1} are jointly Gaussian according to the properties of GP, i.e., $\begin{bmatrix} \mathbf{f}_{1\ldots t} \\ f_{t+1} \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{K} & \mathbf{k} \\ \mathbf{k}^T & k(x_{t+1}, x_{t+1}) \end{bmatrix}\right)$, where $k = [k(x_{t+1}, x_1), k(x_{t+1}; x_2) \ldots k(x_{t+1}, x_t)]$. Therefore, the predictive distribution is $P(f_{t+1}|\{\mathbf{x}_{1\ldots t}, \mathbf{f}_{1\ldots t}\}, x_{t+1}) = \mathcal{N}\left(\mu_t(x_{t+1}), \sigma_t^2(x_{t+1})\right)$, where $\mu_t(x_{t+1}) = \mathbf{k}^T\mathbf{K}^{-1}\mathbf{f}_{1\ldots t}$ and $\sigma_t^2(x_{t+1}) = k(x_{t+1}, x_{t+1}) - \mathbf{k}^T\mathbf{K}^{-1}\mathbf{k}$.

To guide the search for optimum efficiency, Bayesian optimization uses an acquisition function $a(x)$, which is an inexpensive function that can be evaluated at a given point corresponding to potentially low values of the objective function. At each iteration, the acquisition function is minimized to determine the location of the next observation. Although many acquisition functions can be interpreted in the framework of Bayesian decision theory, the most popular acquisition function is expected improvement. Supposing that $f_{min} = \min[f]$ is the minimal value of f observed so far, the expected improvement evaluates f at the point that improves upon $u(x) = \max[0, f_{min} - f(x)]$ the most, and its utility function is f_{min}. If $f(x)$ is less than f_{min}, we receive a reward $f_{min} - f(x)$; otherwise, we receive no reward. The expected improvement acquisition function is then the expected utility as a function of x, i.e., $a(x) = E[u(x)|\{\mathbf{x}_{1\ldots t}, \mathbf{f}_{1\ldots t}\}, x]$. The point with the highest expected improvement is selected. The decision represents an automatic trade-off between exploration (where the objective function is very uncertain) and exploitation (where the objective function is expected to be low).

In summary, the procedure of Bayesian optimization is shown as follows (Brochu et al., 2010):

Step 1 Acquire the observations $\{\mathbf{x}_{1\ldots t}, \mathbf{f}_{1\ldots t}\}$ from previous iterations;

Step 2 Search the location of the next observation x_{t+1} by optimizing the acquisition function over GP $x_{t+1} = \arg\max_x a(x)$;

Step 3 Sample the objective function f_{t+1};

Step 4 Update the GP and then return to Step 1.

After sufficient iterations of sampled values of the objective function, Bayesian optimization can obtain the estimated minimum value and a GP approximation of the objective function.

8. Mathematical model to generate the time series data

Consider a discrete-time dynamical network with N nodes including process noise, described by the following state equation,

$$\mathbf{x}[k+1] = \mathbf{F}(\mathbf{x}[k]) + \mathbf{A}.\mathbf{G}(\mathbf{x}[k]) + \mathbf{w}[k]$$

where $\mathbf{x}[k] = x_i[k], i \in 1, 2, \ldots, N$ is the state vector, $\mathbf{F}(\mathbf{x}[k]) = F_i(x_i[k]), i \in 1, 2, \ldots, N$ is the local dynamics, $\mathbf{G}(\mathbf{x}[k]) = G_{ij}(x_i[k], x_j[k]), i$ and $j \in 1, 2, \ldots, N$ is the coupling dynamics, $\mathbf{w}[k]$ is the process noise that is assumed to be drawn from a zero mean multivariate normal distribution $N(0, \mathbf{Q}^N)$, and $\mathbf{A} = A_{ij}, i$ and $j \in 1, 2, \ldots, N$ is the adjacent matrix describing the topology of the network. If there is a link from node j to node i, $A_{ij} \neq 0$; otherwise, $A_{ij} = 0$.

At each step, the observation $\mathbf{y}[k+1]$ of the true state $\mathbf{x}[k+1]$ is made according to the following observation equation:

$$\mathbf{y}[k+1] = \mathbf{x}[k+1] + \mathbf{v}[k+1]$$

where $\mathbf{v}$ is the observation noise, which is assumed to be drawn from a zero-mean multivariate normal distribution $N(0, \mathbf{R}^N)$.

This general model can offer abundant special cases to generate the time series data. For simplicity, we assume that (i) coupling functions of each node are the same, i.e., $G_{ij}(x_i[k], x_j[k]) = G_i(x_i[k])$ is independent of node j, and (ii) local dynamics is constant, i.e., $F_i(x_i[k]) = \mu_i, i \in 1, 2, \ldots, N$. For each case, the process noise and observation noise of each component are both $N(0, 0.05^2)$, i.e., a normal distribution with mean of 0 and standard deviation of 0.05. The initial value is drawn from a uniform distribution between 0 and 1. Steps from 101 to 150 are adopted to check the possible cointegration.

When the system size is two, three cases to construct the cointegrating relationship are discussed:

(i) Unidirectional coupling. We set adjacency matrix $\mathbf{A} = \begin{bmatrix} 0.9 & 0 \\ -1.35 & 0 \end{bmatrix}$ and local dynamics $\mu = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, and then we discuss three subcases: (A) period oscillation. $G_1(x_1[k]) = \mathrm{Sin}(\frac{\pi k}{5})$ is a sine function; (B) extreme value. $G_1(x_1[k])$ is drawn from a Pareto distribution with a minimum value parameter of 0.1 and a shape parameter of 1.25; (C) chaotic map. $G_1(x_1[k]) = 1 - 2|x_1[k]|$ is the tent map;

(ii) Bidirectional coupling. We set adjacency matrix $\mathbf{A} = \begin{bmatrix} 1/3 & 2/3 \\ 2/9 & 4/9 \end{bmatrix}$ and local dynamics $\mu = \begin{bmatrix} 0 \\ 1/2 \end{bmatrix}$, and then we discuss the same three subcases as before: (A) G_1, G_2 are sine functions; (B) G_1, G_2 are Pareto distributions; (C) G_1, G_2 are tent maps;

(iii) Synchronization coupling. We set adjacency matrix $\mathbf{A} = \begin{bmatrix} 1-\epsilon & \epsilon \\ \epsilon & 1-\epsilon \end{bmatrix}$, where $\epsilon = 0.1, 0.2, 0.3, 0.4, 0.5$, local dynamics $\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and G_1, G_2 are both tent maps.

When the system size is larger than two, we consider four cases:

(i) Unidirectional coupling. We set adjacency matrix $\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 3/4 & 0 & 0 \\ 5/4 & 0 & 0 \end{bmatrix}$, local dynamics $\mu = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$ and coupling dynamics $G_1(x_1[k]) = 1 - 2|x_1[k]|$ as the tent map, $G_2(x_2[k]) = \mathrm{Sin}(\frac{\pi k}{5})$ is a sine function, and $G_3(x_3[k])$ is drawn from a Pareto distribution with a minimum value parameter of 0.1 and a shape parameter of 1.25. In fact, the results are independent of $G_2(x_2[k]), G_3(x_3[k])$;

(ii) Bidirectional coupling. We set adjacency matrix $\mathbf{A} = \begin{bmatrix} 1/6 & 1/3 & 1/2 \\ 1/8 & 1/4 & 3/8 \\ 1/9 & 2/9 & 1/3 \end{bmatrix}$, local dynamics $\mu = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ and coupling dynamics G_1, G_2, G_3 to be the same as before;

(iii) Synchronization coupling. We set adjacency matrix $\mathbf{A} = \begin{bmatrix} 1-\epsilon & \epsilon & 0 \\ \epsilon & 1-\epsilon & 0 \\ 0 & 0 & 1 \end{bmatrix}$ where $\epsilon = 0.5$, local dynamics $\mu = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ and coupling dynamics G_1, G_2, G_3 are all tent maps;

(iv) Mixing coupling. We set adjacency matrix $\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2/3 & 1/3 & 0 \end{bmatrix}$, local dynamics $\mu = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ and coupling dynamics G_1, G_2, G_3 as the tent map, sine function and Pareto distribution.

9. Datasets for real-world examples

9.1. Checking the relationship between models and observations of global warming

The observation time series is extracted from annual mean temperature anomalies in HadCRUT4 filled-in T2m/SST (http://www-users.york.ac.uk/ kdc3/papers/coverage2013/series.html). All climate models included in CMIP5 are considered in this work. Because the models provided many different realizations, we just select the first one. Adopting the anomalous values for global data is recommended as a means to avoid biases that could result from the use of absolute global mean temperatures (Jones et al., 1999). To guarantee the same scale as the observation time series that is already in anomaly form, each model time series is converted to anomaly form. All the model time series are obtained by downloading from KNMI Climate Explorer (http://climexp.knmi.nl/). We select the time period from 1905 to 2005, covered by all observed and modeled datasets.

9.2. Identifying the possible synchronization in electroencephalographic signals

The EEG signal/time series is obtained from electrodes placed on the left and right frontal cortexes of male adult WAG/Rij rats (a genetic model for human absence epilepsy) (Quiroga et al., 2002; van Luijtelaar, 1997). The length of each data segment was 5 seconds, i.e., 1,000 data points. Case 1 corresponds to a normal EEG, and cases 2 and 3 have obvious spike discharges. Since epilepsy is related to an abnormal synchronization in the brain, spikes are generally considered to be a landmark of epileptic activity.

9.3. Determining the possible leadership of Bitcoin in the cryptocurrency market

The cryptocurrency data are extracted from the website Coin Market Cap (https://coinmarketcap.com/). Based on the market capitalization, the price in US dollars and the volume of trading in the preceding 24 h, we collect daily closing values of seven typical cryptocurrencies: Bitcoin (BTC), Ethereum (ETH), Ripple (XRP), Litecoin (LTC), EOS (EOS), Cardano (ADA) and Stellar (XLM). Bitcoin Cash (BCH) has a large market capitalization, and its trade is active and flowing, but it is just a branch of BTC. Therefore, we do not consider it. The period under examination is from April 1st, 2017, to March 31st, 2018, for a total of 182 observations. This period witnesses the recent up and down of the whole cryptocurrency market.

Supplemental References

Brochu, E., Cora, V.M., De Freitas, N., 2010. A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. arXiv preprint arXiv:1012.2599 .

Engle, R.F., Granger, C.W., 1987. Co-integration and error correction: representation, estimation, and testing. Econometrica 55, 251–276.

Friston, K., 2012. The history of the future of the bayesian brain. NeuroImage 62, 1230–1233.

Granger, C.W., Newbold, P., Econom, J., 2001. Spurious regressions in econometrics. A Companion to Theoretical Econometrics, Blackwell, Oxford , 557–561.

Haines, L.M., Perevozskaya, I., Rosenberger, W.F., 2003. Bayesian optimal designs for phase i clinical trials. Biometrics 59, 591–600.

Hamilton, J.D., 1994. Time series analysis. Princeton University Press, Princeton, NJ.

Johansen, S., 1995. Likelihood-based inference in cointegrated vector autoregressive models. Oxford University Press on Demand.

Jones, P.D., New, M., Parker, D.E., Martin, S., Rigor, I.G., 1999. Surface air temperature and its changes over the past 150 years. Reviews of Geophysics 37, 173–199.

Kwiatkowski, D., Phillips, P.C., Schmidt, P., Shin, Y., 1992. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? Journal of econometrics 54, 159–178.

van Luijtelaar, G., 1997. The WAG/Rij Rat Model of Absence Epilepsy: Ten Years of Research: a Compilation of Papers. Nijmegen University Press.

McCracken, J.M., Weigel, R.S., 2014. Convergent cross-mapping and pairwise asymmetric inference. Physical Review E 90, 062903.

Pearl, J., 2009. Causality. Cambridge university press.

Phillips, P.C., Perron, P., 1988. Testing for a unit root in time series regression. Biometrika 75, 335–346.

Quiroga, R.Q., Kraskov, A., Kreuz, T., Grassberger, P., 2002. Performance of different synchronization measures in real data: a case study on electroencephalographic signals. Physical Review E 65, 041903.

Robert, C., 2014. Machine learning, a probabilistic perspective.

Said, S.E., Dickey, D.A., 1984. Testing for unit roots in autoregressive-moving average models of unknown order. Biometrika 71, 599–607.

Shahriari, B., Swersky, K., Wang, Z., Adams, R.P., De Freitas, N., 2016. Taking the human out of the loop: A review of bayesian optimization. Proceedings of the IEEE 104, 148–175.

Shi, Z., Church, R.M., Meck, W.H., 2013. Bayesian optimization of time perception. Trends in cognitive sciences 17, 556–564.

Snoek, J., Larochelle, H., Adams, R.P., 2012. Practical bayesian optimization of machine learning algorithms, in: Advances in neural information processing systems, pp. 2951–2959.

Sugihara, G., May, R., Ye, H., Hsieh, C.h., Deyle, E., Fogarty, M., Munch, S., 2012. Detecting causality in complex ecosystems. science 338, 496–500.

Takens, F., 1981. Detecting strange attractors in turbulence, in: Dynamical systems and turbulence, Warwick 1980. Springer, pp. 366–381.

Tsay, R.S., 2005. Analysis of financial time series. volume 543. John Wiley & Sons.

Tsonis, A.A., Deyle, E.R., Ye, H., Sugihara, G., 2018. Convergent cross mapping: theory and an example, in: Advances in Nonlinear Geosciences. Springer, pp. 587–600.