Universal Cointegration and Its Applications

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

[1]Department of Environmental Science, Policy, and Management, University of California, Berkeley,
chengyitu@berkeley.edu
[2]College of Geography and Environment, Shandong Normal University
[3]School of Economics and Management, Beijing Information Science & Technology University

Identifying the potential cointegration among time series is a challenging and open problem. 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. It is common in even the simplest nonlinear systems, such as shown in Fig. 1, where time series are generated by the following set of state $\{x_1, x_2\}$ and observation $\{y_1, y_2\}$ equations: $\begin{cases} x_1[k+1] = \mu_1 + A_{11}G_1(x_1[k]) + A_{12}G_2(x_2[k]) + w_1[k] \\ x_2[k+1] = \mu_2 + A_{21}G_1(x_1[k]) + A_{22}G_2(x_2[k]) + w_2[k] \end{cases}$ 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 $\begin{cases} x_1[k+1] = \mu_1 + A_{11}G_1(x_1[k]) + A_{12}G_2(x_2[k]) + w_1[k] \\ x_2[k+1] = \mu_2 + \alpha A_{11}G_1(x_1[k]) + \alpha A_{12}G_2(x_2[k]) + w_2[k] \end{cases}$. 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 $\begin{cases} y_1[k+1] = x_1[k+1] + v_1[k+1] \\ y_2[k+1] = x_2[k+1] + v_2[k+1] \end{cases}$ where $v_1[k+1], v_2[k+1]$ are independent normal distributions. Therefore, the following equation holds:

$$y_2[k+1] = \alpha y_1[k+1] + (\mu_2 - \alpha\mu_1) + (w_2[k] - \alpha w_1[k] + v_1[k+1] + v_1[k+1])$$

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 $\alpha y_1[k+1] + (\mu_2 - \alpha\mu_1)$ due to the noise for short time windows, their difference is stationary if we consider sufficiently long time intervals.

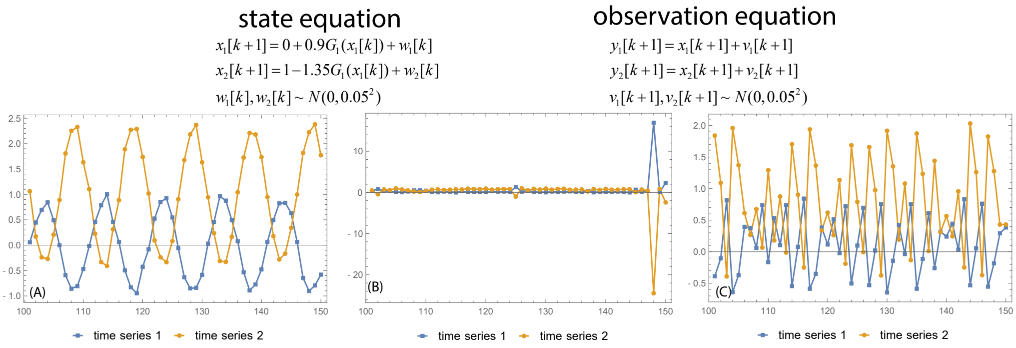

Figure 1: Unidirectional coupling time series for (A) period oscillation where $G_1(x_1[k]) = \text{Sin}(\frac{\pi k}{5})$; (B) extreme value where $G_1(x_1[k])$ is drawn from a Pareto distribution with minimum value parameter of 0.1 and shape parameter of 1.25; (C) chaotic map where $G_1(x_1[k]) = 1 - 2|x_1[k]|$. 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.

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

[1], which is misleading statistical evidence of a linear relationship between independent nonstationary variables. Therefore, applying a regression model to identify long-term stationary relationships is not reliable and valid, 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.

Classic cointegration, such as the Engle-Granger cointegration test [2] and Johansen cointegration test [3], only considers or restricts in cointegration $CI(1,1)$, i.e., (i) all time series $\mathbf{x}[k]$ are nonstationary, and the minimum number of differences required to obtain the stationary series is 1, and (ii) a vector $\mathbf{C} \neq 0$ exists such that $\mathbf{C}^T\mathbf{x}[k]$ is stationary. The prerequisite of them 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 and phase estimation of neurophysiological signals. 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.

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 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 the order of integration is not 1, cases that are not covered by classic cointegration. 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 (a) check the relationship between models and observations of global warming, (b) identify the possible synchronization in electroencephalographic signals and (c) determine the possible leadership of Bitcoin in the cryptocurrency market.

Acknowledgements (optional)

Thank you for spending your time in reading the instructions. This work is under review in iScience. Feel free to contact us for any further information.

References (optional)

[1] Granger C W J, Newbold P, Econom J. Spurious regressions in econometrics[J]. A Companion to Theoretical Econometrics, Blackwell, Oxford, 2001: 557-561.

[2] Engle R F, Granger C W J. Co-integration and error correction: representation, estimation, and testing[J]. Econometrica: journal of the Econometric Society, 1987: 251-276.

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