THE COUPLING SPECTRUM: A NEW METHOD FOR DETECTING TEMPORAL NONLINEAR CAUSALITY IN FINANCIAL TIME SERIES

Amir Reza Alizad-Rahvar – Masoud Ardakani – Ivor Cribben

Abstract

Identifying dynamic causal relationships between financial time series may help explain market dynamics. The Granger causality (G-causality) test is a method to detect linear causal relationships between time series. However, there exists significant evidence for nonlinear causality between financial time series. Hence, several nonlinear extensions of G-causality (NLG-causality) were proposed. Moreover, a new method called the coupling spectrum (CS) was recently proposed to find the nonlinear causal relationship between two time series.

In many financial cases, the direction of causality is changing over time. In this work, we adapt the NLG-causality and CS methods by using a moving window technique to identify possible causality changes over time. We compare the performance of the adapted CS and NLG-causality methods on a simulated temporal nonlinear causal system and a real data set - the stock prices of Apple Inc. and Microsoft Corporation. The simulated and empirical results show that the CS method is more robust than the NLG-causality method and that CS is capable of dealing with time-varying nonlinear causality between financial time series.

Key words: Causality inference, nonlinear causality, Granger causality, time series analysis, stock dynamics

JEL Code: C19, C51, C58

Introduction

The Granger causality (G-causality) test (Granger, 1969) is a statistical hypothesis test for identifying causal relationships between time series. This method estimates a linear regression model with lagged values of the time series $\{d_t\}$ (the driver time series) used to predict the future values of $\{r_t\}$ (the response time series) in the presence of lagged values of $\{r_t\}$. If the error of prediction is reduced by inclusion of $\{d_t\}$, $\{d_t\}$ is the Granger-cause of $\{r_t\}$.

The assumption of linearity in G-causality test can be violated in real applications and it cannot detect nonlinear causal relationships (Brock, 1991). Many investigations in the literature provide evidence of linear and nonlinear causality between financial time series

(Hiemstra & Jones, 1994; Yörük, 2006). Hence, different nonlinear extensions of G-causality (NLG-causality) were proposed to detect nonlinear causality in financial data (Hiemstra & Jones, 1994; Diks & Panchenko, 2006; Dhamala et al., 2006; Papadimitriou et al., 2003).
Recently, a statistical method called the coupling spectrum (CS) (Alizad-Rahvar & Ardakani, 2012) was proposed to detect nonlinear causality between two time series. This nonparametric method can identify causality in different scenarios including unidirectional and bidirectional causality, linear and nonlinear causality, and time series with small and large sample sizes. In this work, we use the CS method to identify causality between financial time series.
In many financial data sets, the direction of causality changes over time. To deal with temporal causality, causality inference methods can be combined with moving window techniques to identify possible causality changes over time. In this paper, we extend the CS method by using a moving window technique and compare its performance on a simulated temporal nonlinear causal system to a moving window adaptation of the NLG-causality method proposed in Hiemstra & Jones (1994). We also compare their performance on a real data set -the stock prices of Apple Inc. and Microsoft Corporation. The simulated and empirical results show that the CS method is more robust than NLG-causality method.
The rest of the paper is organized as follows: in Section 1, we begin by introducing some notation and by setting the framework for the NLG-causality and CS methods. The results from the simulation and real data examples are presented in Section 2. Finally, we conclude with a discussion.

1 Background

1.1 Notation

Consider a cause-effect relationship as a coupled system consisting of a driver system D and a response system R, denoted by $D \rightarrow R$. The samples of D are denoted by a time series $\{d_t\}$, consisting of n time points. Now, define $\mathbf{D}_{t-1} = (d_{t-1}, d_{t-2}, ..., d_{t-L_d})$ representing L_d lagged values of $\{d_t\}$. Similarly, for $\{r_t\}$ we can define $\mathbf{R}_{t-1}$ with the length of lagged values L_r.
Using the maximum norm, we define the distance of the points corresponding to times t and t'

$$\rho_{tt'}^{D} = \left\| \mathbf{D}_{t-1} - \mathbf{D}_{t'-1} \right\| = \max_{1 \le \ell \le L_d} \left| d_{t-\ell} - d_{t'-\ell} \right| \;; \tag{1.a}$$

$$\rho_{tt'}^{R} = \left\| \mathbf{R}_{t-1} - \mathbf{R}_{t'-1} \right\| = \max_{1 \le \ell \le L_r} \left| r_{t-\ell} - r_{t'-\ell} \right| \;; \tag{1.b}$$

$$\rho_{tt'}^{r} = \left| r_t - r_{t'} \right|. \tag{1.c}$$

1.2 Causality and closeness

In this section, we explain the common underpinnings of both the NLG-causality and CS methods. If we have $D \to R$, r_t should be predictable by the lagged values of $\{d_t\}$ and $\{r_t\}$ (for sufficiently large values of L_d and L_r), i.e.,

$$r_t = f(\mathbf{R}_{t-1}, \mathbf{D}_{t-1}). \tag{2}$$

Therefore, provided that $D \to R$, the closeness of the points $\mathbf{D}_{t-1}$ and $\mathbf{D}_{t'-1}$ in the driver space and $\mathbf{R}_{t-1}$ and $\mathbf{R}_{t'-1}$ in the response space imply the closeness of r_t and $r_{t'}$. To quantify the dependency between the closeness of the points in driver and response spaces, we can define a conditional probability based on the distance of the points. If the distance of $\mathbf{R}_{t-1}$ from $\mathbf{R}_{t'-1}$ is smaller than $\delta^r > 0$ (i.e., $\rho^R_{tt'} < \delta^r$), and provided that the distance between $\mathbf{D}_{t-1}$ and $\mathbf{D}_{t'-1}$ is smaller than $\delta^d > 0$ (i.e., $\rho^D_{tt'} < \delta^d$), then the probability that the distance between r_t and $r_{t'}$ is smaller than $\varepsilon^r > 0$ (i.e., $\rho^r_{tt'} < \varepsilon^r$) is denoted by

$$P(\varepsilon^r \mid \delta^r, \delta^d) = P(\rho^r_{tt'} < \varepsilon^r \mid \rho^R_{tt'} < \delta^r, \rho^D_{tt'} < \delta^d) . \tag{3}$$

Both the NLG-causality method (Hiemstra & Jones, 1994) and the CS method are based on the calculation of $P(\varepsilon^r \mid \delta^r, \delta^d)$.

1.3 Nonlinear Granger causality

The Hiemstra-Jones (HJ) method (Hiemstra & Jones, 1994) is a nonlinear extension of G-causality method. The HJ test is a hypothesis test for the following hypothesis

H_0: D does not Granger cause R.

If we define $P_\varepsilon(\varepsilon^r \mid \delta^r, \delta^d) = P(\varepsilon^r = \varepsilon \mid \delta^r = \varepsilon, \delta^d = \varepsilon),$ the HJ test states that the null hypothesis H_0 is true if we have for all $\varepsilon > 0$

$$P_\varepsilon(\varepsilon^r \mid \delta^r, \delta^d) = P_\varepsilon(\varepsilon^r \mid \delta^r). \tag{4}$$

In other words, if D does not cause R, the distance of the points in the response space is independent of the corresponding distances in the driver space.

Under the assumption that $\{d_t\}$ and $\{r_t\}$ are strictly stationary, Hiemstra and Jones introduce the test statistic TVAL and its asymptotic distribution under the null hypothesis H_0

$$\text{TVAL} = \sqrt{n}\ \left(P_\varepsilon(\varepsilon^r \mid \delta^r, \delta^d) - P_\varepsilon(\varepsilon^r \mid \delta^r) \right) \overset{a}{\sim} N(0, \sigma^2(L_d, L_r, \varepsilon)) \tag{5}$$

where the variance of the normal distribution and its estimated value is presented in their appendix. By using the observed value of TVAL, we can make a conclusion about H_0.

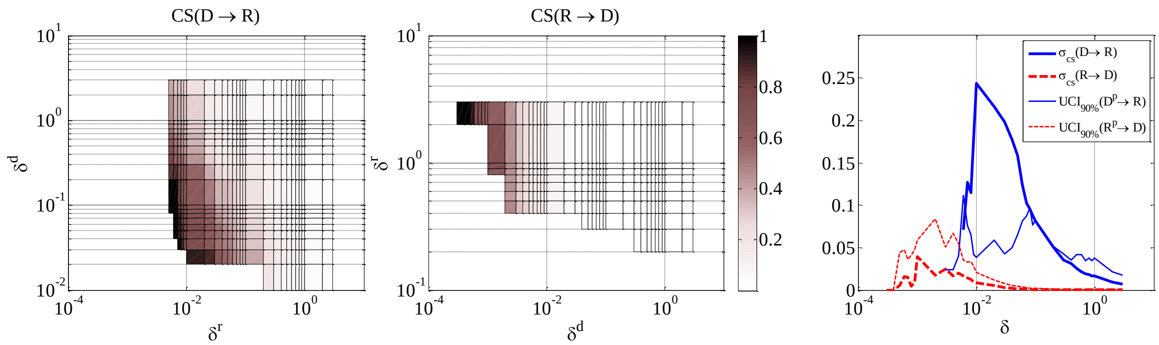
1.4 Coupling spectrum (CS) method

As it is mentioned in Section 1.2, provided that $D \rightarrow R$, the closeness of the points in the driver space implies the closeness of the points in the response space. Therefore, according to equation (2), by increasing $\rho^D_{tt'}$, the probability that r_t stays in the ε^r neighborhood of $r_{t'}$ reduces. Hence, for fixed values of ε^r_o and δ^r_o, $P(\varepsilon^r_o \mid \delta^r_o, \delta^d)$ decreases monotonically as δ^d increases. On the other hand, provided that D does not cause R, denoted by $D \nrightarrow R$, $P(\varepsilon^r_o \mid \delta^r_o, \delta^d)$ does not vary by δ^d, i.e., $P(\varepsilon^r_o \mid \delta^r_o, \delta^d) = P(\varepsilon^r_o \mid \delta^r_o)$. Therefore, by investigating the changes of $P(\varepsilon^r_o \mid \delta^r_o, \delta^d)$ with δ^d, we can detect the causal relationship $D \rightarrow R$.

Consider a causal relationship $D \rightarrow R$ simulated by equation (7) in Section 2, where X and Y represent D and R systems. Figs. 1(a) and 1(b) visualize $P(\varepsilon^r_o \mid \delta^r, \delta^d)$ by a color map for each pair of (δ^r, δ^d). This representation is called the coupling spectrum (CS), denoted by CS($D \rightarrow R$). If we observe a change of color in each column of the CS, this means that $D \rightarrow R$ exists. Otherwise, if all the columns of the CS lack the color change, we can conclude $D \nrightarrow R$. The standard deviation of $P(\varepsilon^r_o \mid \delta^r_o, \delta^d)$ for different values of δ^d, denoted by σ_{cs}, can be used to measure the changes of $P(\varepsilon^r_o \mid \delta^r_o, \delta^d)$ with δ^d in each column of CS($D \rightarrow R$). Fig. 1(c) depicts σ_{cs} corresponding to Figs. 1(a) and 1(b) where $\sigma_{\text{cs}}(D \rightarrow R)$ is considerably greater than $\sigma_{\text{cs}}(R \rightarrow D)$.

To evaluate the significance of σ_{cs} for each direction of causality individually, we permute the driver time series to destroy any dynamical causality and denote the permuted time series by $\{d^p_t\}$. By using the percentile bootstrap method (Efron, 1982), we obtain the $\alpha\%$ confidence interval ($\text{CI}_{\alpha\%}$) of $\sigma_{\text{cs}}(D^p \rightarrow R)$. As an example, Fig. 1(c) shows the upper bound of the

Fig. 1: The coupling spectrum (CS). (a) $D \rightarrow R$: the color of each column changes with δ^d; (b) $R \nrightarrow D$: the color of each column is fixed; (c) σ_{cs} and $\text{UCI}_{90\%}$.



(a) (b) (c) Source: own

$\mathrm{CI}_{90\%}$ ($\mathrm{UCI}_{90\%}$) for both directions of $D \rightarrow R$ and $R \rightarrow D$. As $\sigma_{\mathrm{cs}}(D \rightarrow R)$ lies outside the $\mathrm{UCI}_{90\%}(D^p \rightarrow R)$ for some values of δ^r, we confirm that $\sigma_{\mathrm{cs}}(D \rightarrow R)$ is significant, i.e., D causes R. Conversely, for all values of δ^d, $\sigma_{\mathrm{cs}}(R \rightarrow D)$ is below $\mathrm{UCI}_{90\%}(R^p \rightarrow D)$; hence, we conclude $D \nrightarrow R$.

To evaluate the significance of $\sigma_{\mathrm{cs}}(D \rightarrow R)$ relative to $\mathrm{UCI}_{90\%}(D^p \rightarrow R)$, we use the following measure

$$\mathrm{SIG}_{\mathrm{cs}}(D \rightarrow R) = \sum_{\delta^r} \left[\sigma_{cs}(\delta^r) - \mathrm{UCI}_{90\%}(\delta^r) \right] \times \mathrm{I}\ \ \sigma_{cs}(\delta^r) - \mathrm{UCI}_{90\%}(\delta^r) \tag{6}$$

where $\mathrm{I}(\cdot)$ is an indicator function and $\mathrm{I}(x > 0) = 1;\ \mathrm{I}(x \leq 0) = 0.$ If $D \rightarrow R$, for some values of δ^r, $\sigma_{\mathrm{cs}}(D \rightarrow R)$ is greater than $\mathrm{UCI}_{90\%}(D^p \rightarrow R)$; therefore, $\mathrm{SIG}_{\mathrm{cs}}(D \rightarrow R) > 0$. On the other hand, provided that $D \nrightarrow R$, $\sigma_{\mathrm{cs}}(D \rightarrow R)$ will be smaller than $\mathrm{UCI}_{90\%}(D^p \rightarrow R)$ for all δ^r; consequently, $\mathrm{SIG}_{\mathrm{cs}}(D \rightarrow R) = 0.$

It is noteworthy that $P_\varepsilon(\varepsilon^r \mid \delta^r, \delta^d)$ used in the HJ method is a specific case of $P(\varepsilon^r \mid \delta^r, \delta^d)$ in the CS method where $\varepsilon^r = \delta^r = \delta^d = \varepsilon$. In other words, the HJ method considers the coupling spectrums in Fig. 1 only for one pair of $(\delta^r, \delta^d) = (\varepsilon, \varepsilon)$ and ε^r should also be equal to ε. As we see in Section 2.1, the value of ε has a severe impact on the results of the HJ method. However, in the CS method, we investigate $P(\varepsilon^r \mid \delta^r, \delta^d)$ for the whole range of δ^r and δ^d values and ε^r is determined independently of δ^r and δ^d (for more details about determining the value of ε^r, see (Alizad-Rahvar & Ardakani, 2012)). In Section 2, we illustrate how the flexibility and generality of the parameters in the CS method makes it more robust than the HJ method.

2 Results and applications

In this section, the goal is to discover causal relationships between two time series where the direction of causality is changing over time. To find the temporal changing causality, we use the moving window technique to detect the direction of causality in a small period of time. Now we compare the HJ method with the CS method for simulated and empirical data.

2.1 Simulation results

We evaluate the performance of the HJ and CS methods on simulated data to detect temporal changing causality. Consider two time series $\{x_t\}$ and $\{y_t\}$ having a causal relationship by the Hénon map (Wiesenfeldt et al., 2001)

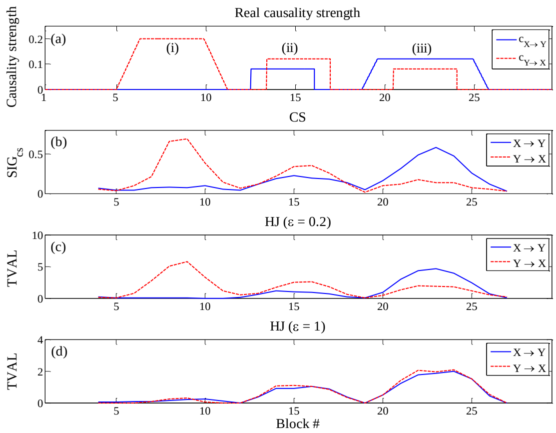
$$x_t = a - x_{t-1}^2 + bx_{t-2} + c_{y\to x}(x_{t-1}^2 - y_{t-1}^2) \quad (7\text{-a})$$

$$y_t = a - y_{t-1}^2 + by_{t-2} + c_{x\to y}(y_{t-1}^2 - x_{t-1}^2) \quad (7\text{-b})$$

where $a = 1.4$, $b = 0.3$ and the initial values of x_0 and y_0 are uniformly distributed in $[0, 0.5]$. Each time series is associated with the squared lagged version of the other one. The strength of causalities between $X \to Y$ and $Y \to X$ are controlled by $c_{x\to y}$ and $c_{y\to x}$, respectively. To have temporal causality in the model, $c_{x\to y}$ and $c_{y\to x}$ change with time as shown in Fig. 2(a). Here, we use the overlapping window with window length N_w. In each step, the window moves $N_f < N_w$ time points further. In the simulation, we used $N_w = 300$ and $N_f = 60$. The lag-lengths L_d and L_r are set to 2. A significant level of 5% is used for the HJ method and we estimate the $\mathrm{UCI}_{90\%}$ for the CS method.

Figure 2 shows the comparison of the CS and HJ methods for 50 trials with different initial values of x_0 and y_0 in equation (7). The mean of the $\mathrm{SIG}_{\mathrm{cs}}$ and TVAL values over the 50 trials are plotted. Figure 2(b) shows that the outcome of the CS method is consistent with the real causal relationships. In other words, the CS method 1) correctly detects the direction of causality in all three parts (the detected causality $X \to Y$ in part (i) is very weak); 2) distinguishes the strong and weak causality in the bidirectional scenarios; 3) finds for each direction of causality the correct ratios of causality strengths in different parts that are proportional to real ratios.

Fig. 2: Finding temporal causality for the simulated data. (a) The real strength of the temporal causality; (b) CS method; (c) and (d) HJ method for different values of ε.



Source: own

Figures 2(c) and 2(d) show the performance of the HJ method for $\varepsilon = 0.2$ and $\varepsilon = 1$, respectively. For $\varepsilon = 0.2$, HJ does not detect any $X \rightarrow Y$ causality in part (i). In parts (ii) and (iii), HJ performs as well as CS. However, Fig. 2(d) illustrates that increasing ε adversely affects the HJ method. In this case, the $Y \rightarrow X$ causality in part (i) is not detected; weak and strong causalities in bidirectional scenarios are not distinguishable; and for each direction of causality, the ratios of causality strengths in different parts are not proportional to the real ratios. Indeed, as HJ considers $P(\varepsilon^r \mid \delta^r, \delta^d)$ just for a specific value of ε, the validity of the HJ results depends severely on ε and in all practical applications ε will be unknown.

2.2 Empirical results

In this section, we investigate the temporal causality between the stock prices of Apple Inc. (AAPL) and Microsoft Corporation (MSFT). A total of 3199 daily stock prices during the time between January 2000 and August 2012 are used. To render each time series weakly stationary, we carry out a piecewise linear detrending. We select a window length of five month and N_f is the duration of one month. The lag-lengths L_d and L_r are set to 5, i.e., we investigate the causal effect of the stock prices of past five business days on the price of the next business day. In addition, a test significant level of 5% and $\varepsilon = 0.7$ are used in the HJ method (this value of ε results in larger TVALs). $UCI_{90\%}$ is estimated for the CS method.

The temporal causalities AAPL$\rightarrow$ MSFT and MSFT$\rightarrow$ AAPL derived by the CS and HJ methods are plotted in Figs. 3 and 4, respectively. The months in these figures represent the middle month of each block. As an evidence for detected causality, the timeline of AAPL and

MSFT major products are depicted by arrows in subplots (a) and (b), respectively. Figures 3 and 4 reveal the following results:

- The direction of causality between these two companies changes over time. Therefore, to investigate causality between financial time series over a long period of time, we have to use a moving window to deal with this time-varying causality.
- Most of the products of each company affect the other one's stock price immediately or a couple of months after each product release. However, the number of the causal relationships detected by the HJ method is less than that of the CS method.
- There are detected causalities that could be due to other factors other than products releases, e.g., detected causalities in the second half-year of 2008 in MSFT $\rightarrow$ AAPL.
- In general, for both methods, it can be concluded that the causal effect of AAPL on MSFT's stock price is greater over time than vice versa.

Fig. 3: Temporal causality between the stock prices of Apple (AAPL) and Microsoft (MSFT) detected by the CS method. (a) AAPL $\rightarrow$ MSFT, (b) MSFT $\rightarrow$ AAPL.

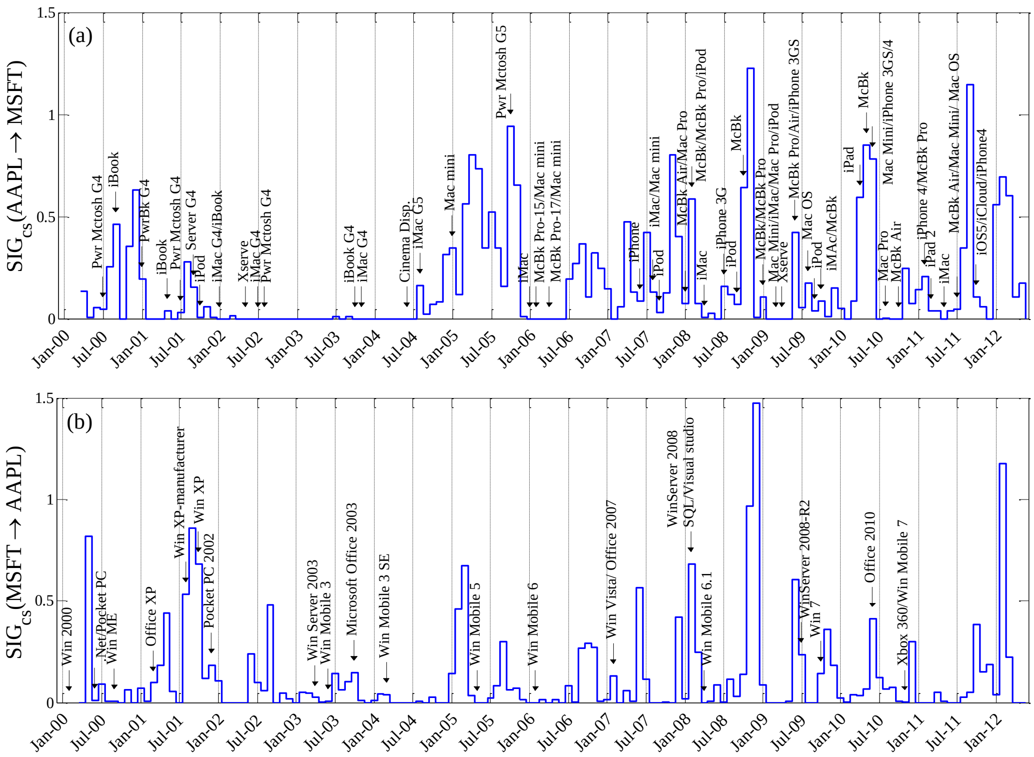
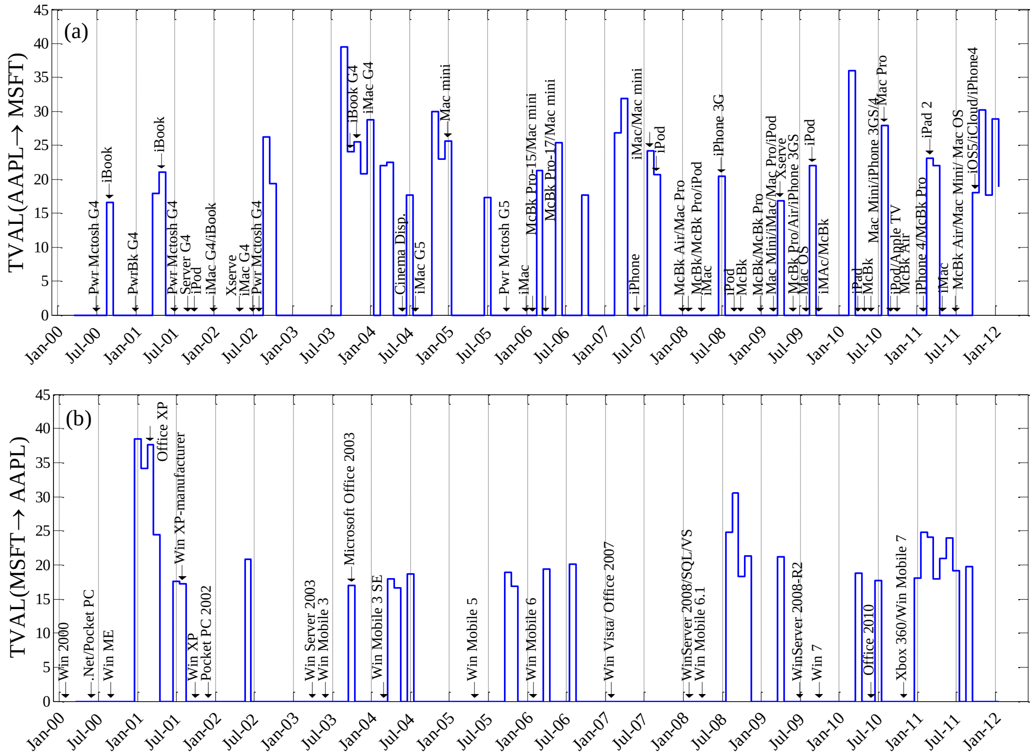


Fig. 4: Temporal causality between the stock prices of AAPL and MSFT derived by the HJ method ($\varepsilon = 0.7$). (a) AAPL $\rightarrow$ MSFT, (b) MSFT $\rightarrow$ AAPL.



Source: own

Conclusion

The dynamic causal relationships between many financial time series have a nonlinear and time-varying nature. In this paper, we extended a recently proposed approach called the coupling spectrum (CS) to detect temporal nonlinear causalities between financial time series. We compared two nonlinear causality inference methods, the HJ and CS methods, and used the overlapping moving window technique to deal with temporal causalities. Examination of these two methods on a simulated nonlinear causal relationship showed that due to the generality of the CS parameters over the HJ parameters, the performance of the CS method is more robust than the HJ method. In other words, HJ can be severely affected by its parameter value selection.

In the final section we applied the CS and HJ methods to the stock prices of two companies, Apple Inc. and Microsoft Corporation, over a decade to detect the temporal causal effects of their stock prices on each other. We found that the direction of causality changes over time, especially around the advent of new products. Hence, in conclusion, in analyzing causality between financial time series over long periods of time, we have to use moving window techniques to deal with the time-varying causality.

References

[1] Alizad-Rahvar, A. R., & Ardakani, M. (2012). Finding weak directional coupling in multiscale time series. *Phys. Rev. E, 86*, 016215.

[2] Brock, W. A. (1991). *Causality, chaos, explanation and prediction in economics and finance*. (pp. 230-279). Boca Raton, FL: CRC Press.

[3] Dhamala, M., Rangarajan, G., & Ding, M. (2006). Estimating Granger causality from Fourier and wavelet transforms of time series data. *Phys. Rev. Lett.*, *100*(1), 018701.

[4] Diks, C., & Panchenko, M. (2006). A new statistic and practical guidelines for nonparametric Granger causality testing. *Journal of Economic Dynamics & Control*, *30*, 1647-1669.

[5] Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*. Philadelphia, PA, USA: SIAM.

[6] Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. Econometrica, 37 (3), 424–438.

[7] Hiemstra, C. & Jones, J. D. (1994). Testing for linear and nonlinear Granger causality in the stock price-volume relation. *Journal of Finance, 49*, 1639–1664.

[8] Papadimitriou, S., Brockwell, A., & Faloutsos, C. (2003). *Adaptive, hands-off stream mining*. Vldb, Berlin, Germany.

[9] Wiesenfeldt, M., Parlitz, U., & Lauterborn, W. (2001). Mixed state analysis of multivariate time series. *Int. J. Bifurcation Chaos, 11*(8), 2217-2226.

[10] Yörük, N. (2006). Testing for linear and nonlinear Granger causality in the stock price-volume relation: Turkish banking firms' evidence. *Applied Financial Economics Letters*, *2*(3), 165-171.

Contact

Amir Reza Alizad-Rahvar

Department of Electrical and Computer Engineering, University of Alberta

2nd floor, ECERF Building, University of Alberta, Edmonton, AB, Canada, T6G 2V4

ar.alizad@ualberta.ca

Masoud Ardakani

Department of Electrical and Computer Engineering, University of Alberta,

2nd floor, ECERF Building, University of Alberta, Edmonton, AB, Canada, T6G 2V4

ardakani@ualberta.ca

Ivor Cribben

Department of Finance and Statistical Analysis, Alberta School of Business, University of Alberta

2-32G Business Building, University of Alberta, Edmonton, AB, Canada, T6G 2R6

cribben@ualberta.cabagoom