STATISTICS IN TRANSITION new series, June 2021
Vol. 22, No. 2, pp. 1–14, DOI 10.21307/stattrans-2021-014
Received – 07.04.2021; accepted – 26.05.2021

Extended residual coherence with a financial application

Xuze Zhang,[1] **Benjamin Kedem**[2]

ABSTRACT

Residual coherence is a graphical tool for selecting potential second-order interaction terms as functions of a single time series and its lags. This paper extends the notion of residual coherence to account for interaction terms of multiple time series. Moreover, an alternative criterion, integrated spectrum, is proposed to facilitate this graphical selection. A financial market application shows that new insights can be gained regarding implied market volatility.

Key words: interaction, residual coherence, nonlinear, time series, volatility index.

1. Introduction

Nonlinear phenomena in random processes have attracted much attention going back to the work of Wiener (1958) concerning random nonlinear oscillators excited by a random input, random shot effect as input for testing nonlinear circuits, and more generally concerning a class of nonlinear polynomial functionals to model input-output relationships in nonlinear systems. In Wiener's words, he was interested in "methods of handling the spectrum," which motivates the use of higher order spectra dealt with by quite a few authors including Brillinger (1965), Brillinger and Rosenblatt (1967), Hinich (1979), Nikias and Mendel (1993), and Elgar et al. (1998). The excellent review paper by Sanaullah (2013) provides numerous additional references about applications of nonlinear techniques based on higher order spectra. Inherent in all nonlinear systems is the problem of assessing the degree and extent of nonlinearity, which can be approached by the detection of nonlinear components or *interactions* (Tick (1961), Elgar et al. (1998)).

In this paper, the detection of nonlinear second-order interactions is done by an extension of *residual coherence* introduced in Khan, Katzoff and Kedem (2014) and applied in mortality forecasting. Residual coherence is a nonlinear variation of the well-known measure of linear coherence. The method is then applied to two volatility indices, the Chicago Board Options Exchange Volatility Index (VIX), and the Russell 2000 Volatility Index (RVX).

[1]Department of Mathematics and Institute for Systems Research, University of Maryland. E-mail: xzhang51@umd.edu. ORCID: https://orcid.org/0000-0002-8672-8515.
[2]Department of Mathematics and Institute for Systems Research, University of Maryland. E-mail: bnk@umd.edu. ORCID: https://orcid.org/0000-0002-7945-8713.

2. Extensions of residual coherence

2.1. Preliminaries

The coherence between two time series $(X(t),Y(t))$ measures the extent of linear relationship between them in the frequency domain. Provided all auto- and cross-spectra exist, it is defined as

$$\gamma_{XY}(\lambda) = \frac{|f_{XY}(\lambda)|^2}{f_{XX}(\lambda)f_{YY}(\lambda)}$$

(see Koopmans (1974)) where f_{XX} and f_{YY} are the spectra of $X(t)$ and $Y(t)$, respectively, and f_{XY} is the cross-spectral density of $X(t)$ and $Y(t)$. This is widely used in detecting connections and clustering of time series. Relevant works include Sun, Miller and D'Esposito (2004), Maharaj and D'Urso (2010) and Euan, Sun and Ombao (2019), among many others. When the relationship is nonlinear, it is frequently analyzed by bispectra, trispectra, or higher-order spectra. For example, a bispectral method for detecting lag processes was proposed by Hinich (1979). Lagged coherence and residual coherence were first introduced in Kedem-Kimelfeld (1975) and Khan, Katzoff and Kedem (2014), respectively, to detect and select potential interaction effects as input to nonlinear systems, based on an orthogonal decomposition in Kimelfeld (1974) without involving bispectrum or higher-order spectra.

Let $Y(t)$ be the output of a system of which the input consists of linear and quadratic filters of $X(t)$ plus noise $\varepsilon(t)$,

$$Y(t) = L[X(t)] + \sum_{k=1}^{\infty} L_{u_k}[\tilde{X}_{u_k}(t)] + \varepsilon(t)$$

where $\tilde{X}_{u_k}$ is a lag process defined as $\tilde{X}_{u_k}(t) = X(t)X(t-u_k) - E[X(t)X(t-u_k)]$. For simplicity, assume that $Y(t)$ and $X(t)$ are zero-mean real valued jointly stationary processes and that all relevant auto- and cross-spectra exist. Then, for sufficiently large n, $Y(t)$ can be approximated by

$$Y^*(t) = G_1(t) + \sum_{k=1}^{n} G_{2,k}(t) + \varepsilon(t) \tag{1}$$

where $G_1(t)$ is a linear filter of $X(t)$, and as in Kedem-Kimelfeld (1975), $G_{2,k}(t)$ is a sum of a linear filter of $X(t)$ and a linear filter of $\tilde{X}_{u_k}(t)$, such that $G_{2,k}(t) \perp G_1(t)$ for $k = 1,\ldots,n$.

Kedem-Kimelfeld (1975) showed that if there is prior knowledge that $Y(t)$ takes on a simpler form

$$Y(t) = L[X(t)] + L_u[\tilde{X}_u(t)] + \varepsilon(t), \tag{2}$$

then it can be rewritten as a sum of two orthogonal processes, $G_1(t)$ and $G_2(t;u)$ plus noise $\varepsilon(t)$,

$$Y(t) = G_1(t) + G_2(t;u) + \varepsilon(t). \tag{3}$$

Then the lag process, or interaction, $\tilde{X}_u(t)$ that minimizes $E\varepsilon^2(t)$ can be selected by finding the lag u that maximizes the lagged coherence $S_2(\lambda;u)$ over all frequencies $\lambda \in [-\pi,\pi]$

such that

$$S_2(\lambda;u) = \frac{f_{G_1G_1}(\lambda) + f_{G_2G_2}(\lambda;u)}{f_{YY}(\lambda)}.$$

However, as mentioned in Kedem-Kimelfeld (1975), there might not exists a u that maximizes $S_2(\lambda;u)$ over all frequencies. One way to resolve this issue is to define the residual coherence as

$$RC(u) = \sup_{\lambda} \frac{f_{G_2G_2}(\lambda;u)}{f_{YY}(\lambda)}$$

and find u that maximizes $RC(u)$. This is shown to be useful for interaction selection in Khan, Katzoff and Kedem (2014) and Kedem (2016).

2.2. Lagged coherence and residual coherence for more than two orthogonal components

Consider the model

$$Y(t) = \sum_{k=1}^{n} L_{k,u_k}[X_{k,u_k}(t)] + \varepsilon(t). \tag{4}$$

The goal is to select $X_{k,u_k}(t)$ from a certain family of processes $\{X_{k,u_k}(t) : u_k = 1,2,\dots\}$ for $k = 1,\dots,n$. Assume that all relevant series are jointly stationary and all relevant auto- and cross-spectra exist. This reduces to (2) when $n = 2$ and $L_{1,u_1}[X_{1,u_1}(t)]$ is the linear filter of $X(t)$. For $n > 2$, we shall extend the orthogonal decomposition (3)

$$Y(t) = \sum_{k=1}^{n} G_k(t;u_1,\dots,u_k) + \varepsilon(t)$$

where all G_k's for $k = 1,\dots,n$ are mutually orthogonal, to account for more orthogonal components, given by

$$G_k(t;u_1,\dots,u_k) = \sum_{j=1}^{k} \int_{-\pi}^{\pi} e^{it\lambda} A_{j,k-j+1}(\lambda) dZ_{X_{j,u_j}}(\lambda)$$

for $k = 1,\dots,n$, where the A's are non-zero and Z's are the corresponding spectral measures.

The A's can be obtained by using the orthogonal conditions among G_k's such that

$$A_{k,1}(\lambda) = \overline{\left[\frac{\sum_{j=1}^{k} c_{k,j}(\lambda) f_{X_{j,u_j}Y}(\lambda)}{\sum_{j=1}^{k} c_{k,j}(\lambda) f_{X_{j,u_j}X_{k,u_k}}(\lambda)} \right]} \qquad k = 1,\dots,n$$

and

$$A_{j,k-j+1}(\lambda) = c_{k,j}(\lambda) A_{k,1}(\lambda) \qquad j = 1,\dots,k, \qquad k = 1,\dots,n$$

where

$$c_{k,k}(\lambda) = 1, c_{k,j}(\lambda) = \frac{\boldsymbol{F}_{k,j}(\lambda)}{\boldsymbol{F}_k(\lambda)}, \boldsymbol{F}_k(\lambda) = (f_{i,j}(\lambda))_{(k-1)\times(k-1)},$$

$f_{i,j} \equiv f_{X_{j,u_j} X_{i,u_i}}$ and $\boldsymbol{F}_{k,j}(\lambda)$ is equivalent to $\boldsymbol{F}_k(\lambda)$, of which jth column is replaced by

$$\boldsymbol{f}_k(\lambda) \equiv -[f_{1,k}(\lambda), \ldots, f_{k-1,k}(\lambda)]^T.$$

More details are provided in Appendix.

Subsequently,

$$\begin{aligned} f_{G_k G_k}(\lambda; u_1, \ldots, u_k) &= |A_{k,1}(\lambda)|^2 \left[\sum_{j=1}^{k} c_{k,j}(\lambda) f_{X_{j,u_j} X_{k,u_k}}(\lambda) \right] \\ S_k(\lambda; u_1, \ldots, u_k) &= \frac{\sum_{j=1}^{k} f_{G_j G_j}(\lambda; u_1, \ldots, u_j)}{f_{YY}(\lambda)} \\ RC(u_1, \ldots, u_k) &= \sup_{\lambda} [S_k(\lambda; u_1, \ldots, u_k) - S_m(\lambda; u_1, \ldots, u_m)] \quad 1 \le m \le k \end{aligned} \tag{5}$$

for $k = 1, \ldots, n$. Note that $S_k(\lambda; u_1, \ldots, u_k)$ and $RC(u_1, \ldots, u_k)$ depend only on $u_{m+1}, \ldots, u_k$ once $u_1, \ldots, u_m$ are determined. The estimates of the above quantities are obtained based on the estimates of the relevant auto- and cross-spectra. Also, to avoid confusion, we denoted the residual coherence in (5) by $RC_{(m+1):k}(u_{m+1}, \ldots, u_k)$ when $u_1, \ldots, u_m$ are determined.

2.3. Selection Criteria

In this section, lagged coherence and residual coherence are examined and an alternative criterion is proposed. Take $n = 2$ and fix u_1, then it reduces to the case in Kedem-Kimelfeld (1975). It illustrates that if there exists a u_2 that maximizes $S_2(\lambda; u_2)$ for all λ, then such u_2 minimizes $E\varepsilon^2(t)$ in

$$E\varepsilon^2(t) = \int_{-\pi}^{\pi} f_{\varepsilon\varepsilon}(\lambda) d\lambda = \int_{-\pi}^{\pi} f_{YY}(\lambda)[1 - S_2(\lambda; u_2)] d\lambda.$$

Indeed, the quantity we wish to maximize is $\int_{-\pi}^{\pi} f_{G_2 G_2}(\lambda; u_2) d\lambda$ since

$$\int_{-\pi}^{\pi} f_{YY}(\lambda) S_2(\lambda; u_2) d\lambda = \int_{-\pi}^{\pi} [f_{G_1 G_1}(\lambda) + f_{G_2 G_2}(\lambda; u_2)] d\lambda$$

based on (5). Such criterion works even if such u_2 does not exist so that this can be an alternative to residual coherence. This criterion can be readily extended to a more general case. Suppose there is prior knowledge for the inclusion of first m processes, i.e. $u_1, \ldots, u_m$ are fixed, then we define the integrated spectrum

$$IS_{(m+1):n}(u_{m+1}, \ldots, u_n) \equiv \int_{-\pi}^{\pi} \sum_{k=m+1}^{n} f_{G_k G_k}(\lambda; u_k, \ldots, u_n) d\lambda$$

and find $u_{m+1}, \ldots, u_n$ that maximizes $IS_{(m+1):n}(u_{m+1}, \ldots, u_n)$.

Once all u's are determined, the regression-based selection method proposed in Khan, Katzoff and Kedem (2014) and Kedem (2016) is used to select significant terms within the processes selected by the graphical method and this is illustrated in both Section 3 and

Section 4. Relevant regression for time series that the selection entails can be found in Kedem and Fokianos (2002).

3. Simulation

In this section, a simulation is performed with $n = 4$ and $u1, u2$ fixed to validate and compare the two criteria, residual coherence and integrated spectrum. The steps are as follows:

1. Generate $\{x_1(t)\}_{t=1}^{1010}$ from an AR(1) process $X_1(t) = 0.4X_1(t-1) + u_1(t)$ and $\{x_2(t)\}_{t=1}^{1010}$ from an AR(1) process $X_2(t) = 0.2X_2(t-1) + u_2(t)$, where u's are white noise $N(0,1)$.

2. Obtain

$$y(t) = 0.4x_1(t) + 0.3x_2(t) + 0.4x_1(t-2)x_2(t-1) + 0.3x_1(t)x_2(t-4) + \varepsilon(t), \quad (6)$$

   where ε's are white noise $N(0,1)$ and t=11,...,1010 so that all relevant series have length 1000.

3. This is the model (4) with $n = 4$ and known $X_1(t)$, $X_2(t)$. We considered selecting $X_{3,u_3}(t)$ and $X_{4,u_4}(t)$ from the family $\{X_1(t+h)X_2(t) : h = -9,-8,\ldots,0,\ldots,9\}$. In fact, this family can be made larger and the choice here only serves as an example. Then, we estimated all relevant auto- and cross-spectra using Tukey-Hamming kernel with window size 10 for frequencies $\lambda_k = -\pi + k\pi/1000$, $k = 0,\ldots,2000$. Subsequently, we estimated $RC_{3:3}(u_3)$, $IS_{3:3}(u_3)$, $RC_{4:4}(u_4)$ and $IS_{4:4}(u_4)$ for $u_3, u_4 = -9,-8,\ldots,0,\ldots,9$. Note that

$$\widehat{IS}_{3:3}(u_3) = \sum_{k=1}^{2000} \pi \hat{f}_{G_3G_3}(\lambda_k; u_3)/1000$$
$$\widehat{IS}_{4:4}(u_4) = \sum_{k=1}^{2000} \pi \hat{f}_{G_4G_4}(\lambda_k; u_3, u_4)/1000$$

   and $RC_{4:4}(u_4)$, $IS_{4:4}(u_4)$ only depend on u_4 once u_3 is fixed.

The results are shown by Figure 1, which indicates that the process $X_1(t-1)X_2(t)$ is the optimal choice for the third input. It is also observed from Figure 1 that $X_1(t+4)X_2(t)$ is another potential input since the bars that correspond to $u_3 = 4$ are the second highest ones in both graphs. With $u_3 = -1$ fixed, u_4 can be determined by $\widehat{RC}_{4:4}(u_4)$ and $\widehat{IS}_{4:4}(u_4)$, as shown by Figure 2, and both graphs indicate that $u_4 = 4$ is the optimal choice, which accords with the original model (6).

With u_3 and u_4 determined, we select significant covariates from the selected processes $X_1(t-1)X_2(t)$ and $X_1(t+4)X_2(t)$ using the regression-based method in Khan, Katzoff and Kedem (2014) and Kedem (2016). We selected four lag terms from each input, i.e. $x_1(t),\ldots,x_1(t-3),x_2(t),\ldots,x_2(t-3),x_1(t-1)x_2(t),\ldots,x_1(t-4)x_2(t-3),x_1(t)x_2(t-4),\ldots,x_1(t-3)x_2(t-7)$ and regressed $y(t)$ on all the selected covariates. We then performed stepwise selection based on AIC and it is observed from Table 1 that the selected

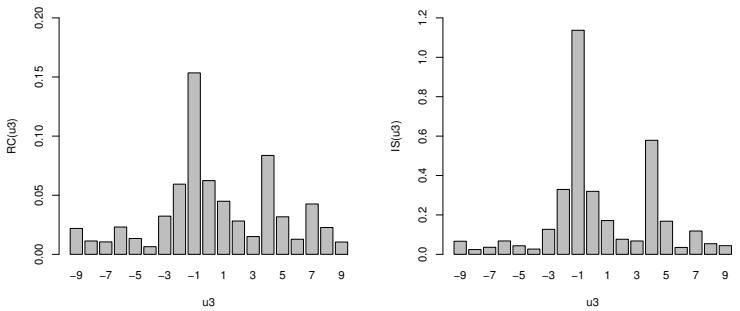

Figure 1. $\widehat{RC}_{3:3}(u_3)$ (left) and $\widehat{IS}_{3:3}(u_3)$ (right) for $u_3 = -9, \ldots, 9$.

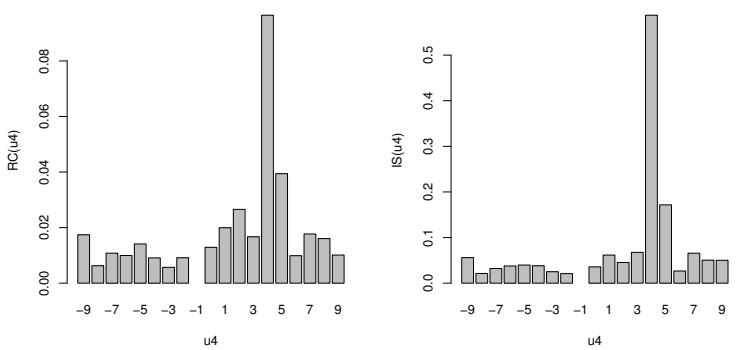

Figure 2. $\widehat{RC}_{4:4}(u_4)$ (left) and $\widehat{IS}_{4:4}(u_4)$ (right) for $u_4 = -9, \ldots, 9$. Note that the bars that correspond to $u_4 = -1$ are set to be 0 since $u_3 = -1$.

Table 1. Regression result of the selected model

	Estimate	SE	p-value
Intercept	-0.0264	0.0322	0.4125
$x_1(t)$	0.3876	0.0292	0.0000
$x_2(t)$	0.2907	0.0303	0.0000
$x_2(t-2)$	0.0571	0.0307	0.0629
$x_2(t-3)$	-0.0555	0.0307	0.0708
$x_1(t-2)x_2(t-1)$	0.3729	0.0277	0.0000
$x_1(t)x_2(t-4)$	0.2569	0.0275	0.0000

model is similar to (6) since the significant ($\alpha = 0.05$) covariates are identical to the ones in (6) and the estimated coefficients correspond to the ones in (6), which validates the method.

Remark: It seems that regression-based selection criteria of interaction terms can be applied directly, thus bypassing the need for our graphical method. However, we rationalize the use of our spectral graphical selection for the following reason. The number of potential covariates in the initial model might be too large, which could result in conflicting selections and possible inconsistencies depending on the model selection method. Our graphical method identifies potentially useful interactions which can then be taken into account and reduce significantly the number of covariates fed into any model selection method, thus rendering the selection more manageable.

4. An Application to Volatility Index

The Volatility Index of a certain underlying asset gives the expectation of the corresponding market volatility in a certain future period. The first and most famous one, VIX, was introduced by Whaley (1993). The underlying asset for VIX is the S&P 500 index so that it reflects the implied volatility of the stock performance of large capitalization companies. For the implied volatility of small capitalization stocks, we have chosen RVX. These two volatility indices shall be considered here as indicators for the stock market. For commodity markets, two important volatility indices, the Crude Oil Exchange Traded Funds Volatility Index (OVX) and the Gold Exchange Traded Funds Volatility Index (GVX) were used. The two-year (2018-2019) daily data of these four series were taken from the Federal Reserve Economic Data Website (https://fred.stlouisfed.org/).

The above methods were applied to analyze the relationships between the volatility indices of the stock and commodity markets. This section is divided into two parts, one investigates the influence of OVX and GVX on VIX and the other examines the influence on RVX.

Before the analysis, the four series were pre-processed to render them approximately stationary. That was achieved by first-order differencing of the original series and centring at zero. Figure 3 depicts the four series before and after processing.

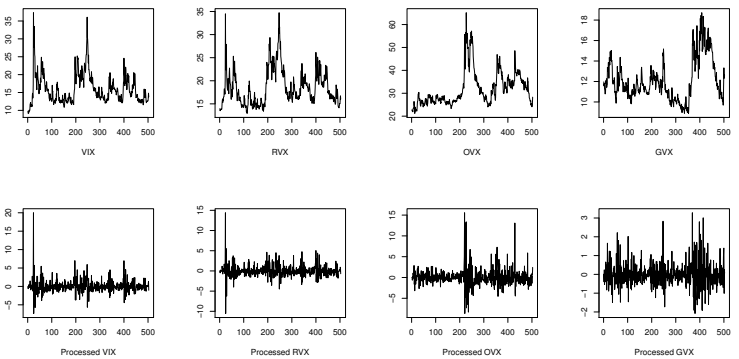

Figure 3. VIX, RVX, OVX and GVX series before and after processing

Table 2. Regression result of $y(t)$ on $x_1(t)$ and $x_2(t)$

	Estimate	SE	p-value
$x_1(t)$	0.1486	0.0347	0.0000
$x_2(t)$	0.9395	0.1075	0.0000

4.1. VIX, OVX and GVX

Consider the processed VIX series as the output and the processed OVX and GVX series as the input. Denote the processed VIX as $y(t)$, processed OVX as $x_1(t)$, and processed GVX as $x_2(t)$. The results of linear regression of $Y(t)$ on $x_1(t)$ and $x_2(t)$ in Table 2 indicate that it is reasonable to include these two series as input since their coefficients are significant. Note that the intercept is omitted since all three series were centred at zero.

Then, the goal is to find the third input based on the cross products of $x_1(t)$ and $x_2(t)$. This resembles the simulation problem so that we perform the same analysis as we did in Section 3. We first select the third input from the family of processes $\{X_1(t+h)X_2(t) : h = -9,-8,\ldots,0,\ldots,9\}$ and the estimated *RC*'s and *IS*'s are shown in Figure 4 and it is observed that both criteria indicate that $u_3 = 4$ is the optimal choice. The u_4 is checked with $u_3 = 4$ fixed and Figure 5 shows that none of the bars is particularly prominent so that we stop at the third input. In correspondence to Section 3, we selected four lag terms from each of the three input series and performed a stepwise selection. The final model selected by AIC is shown in Table 3. It includes the two significant (α=0.05) interaction terms $x_1(t)x_2(t-4)$ and $x_1(t-1)x_2(t-5)$.

4.2. RVX, OVX and GVX

We repeated the analysis in Section 4.1 with VIX replaced by RVX. We still consider the processed OVX and GVX as input and try to detect possible significant interactions. In

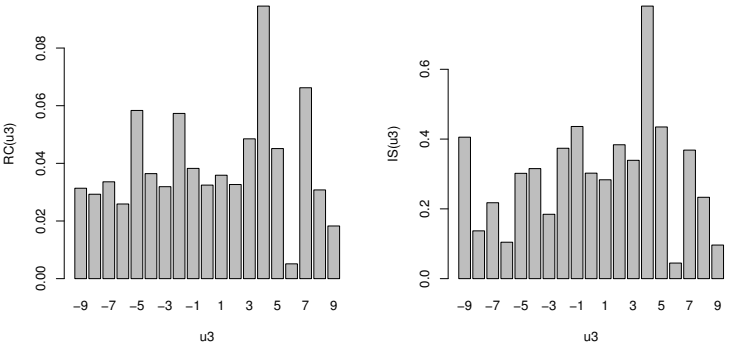

Figure 4. $\widehat{RC}_{3:3}(u_3)$ (left) and $\widehat{IS}_{3:3}(u_3)$ (right) for $u_3 = -9,\ldots,9$.

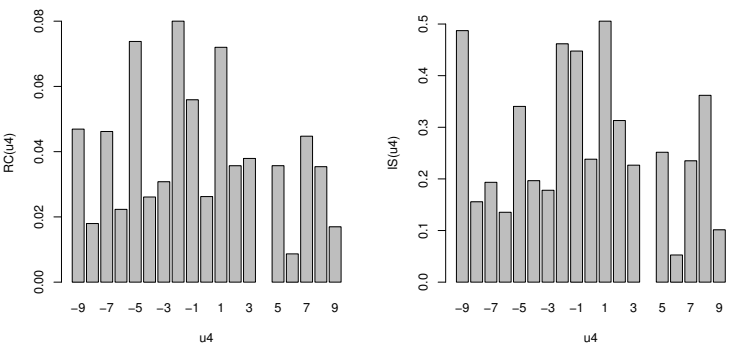

Figure 5. $\widehat{RC}_{4:4}(u_4)$ (left) and $\widehat{IS}_{4:4}(u_4)$ (right) for $u_4 = -9,\ldots,9$. Note that the bars that correspond to $u_4 = 4$ are set to be 0 since $u_3 = 4$

Table 3. Regression result of the selected model

	Estimate	SE	p-value
Intercept	-0.0007	0.0717	0.9927
$x_1(t)$	0.1677	0.0360	0.0000
$x_2(t)$	0.9039	0.1133	0.0000
$x_2(t-1)$	-0.2504	0.1078	0.0206
$x_1(t)x_2(t-4)$	0.1344	0.0498	0.0072
$x_1(t-1)x_2(t-5)$	-0.1010	0.0508	0.0473
$x_1(t-2)x_2(t-6)$	-0.0914	0.0504	0.0707
$x_1(t-3)x_2(t-7)$	0.0817	0.0501	0.1037

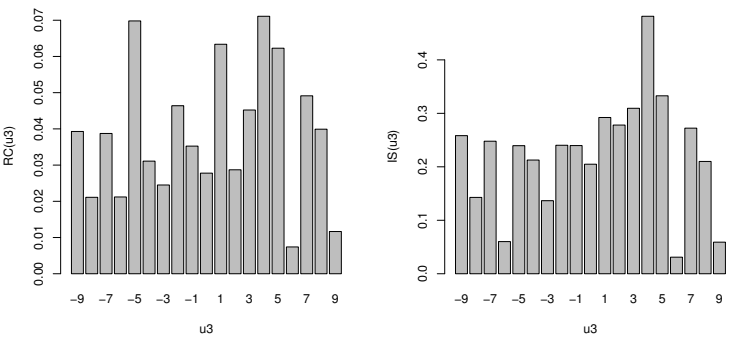

Figure 6. $\widehat{RC}_{3:3}(u_3)$ (left) and $\widehat{IS}_{3:3}(u_3)$ (right) for $u_3 = -9,\ldots,9$.

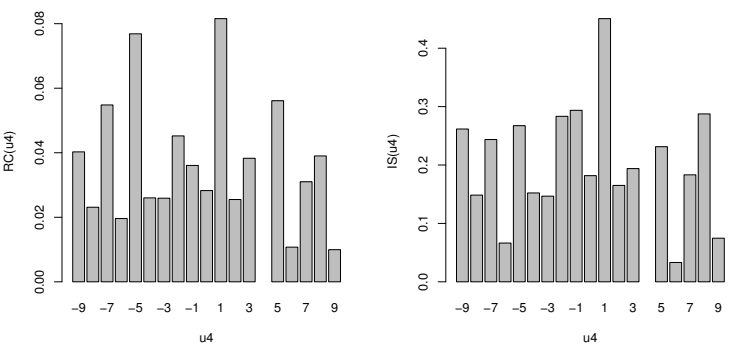

Figure 7. $\widehat{RC}_{4:4}(u_4)$ (left) and $\widehat{IS}_{4:4}(u_4)$ (right) for $u_4 = -9,\ldots,9$. Note that the bars that correspond to $u_4 = 4$ are set to be 0 since $u_3 = 4$

Table 4. Regression result of the selected model

	Estimate	SE	p-value
Intercept	0.0035	0.0626	0.9549
$x_1(t)$	0.1253	0.0314	0.0001
$x_2(t)$	0.8296	0.0994	0.0000
$x_2(t-1)$	-0.2130	0.0954	0.0261
$x_1(t)x_2(t-4)$	0.1191	0.0440	0.0071
$x_1(t-1)x_2(t-5)$	-0.1114	0.0448	0.0132
$x_1(t-3)x_2(t-7)$	0.0632	0.0432	0.1438
$x_1(t)x_2(t-1)$	0.1049	0.0512	0.0410
$x_1(t-1)x_2(t-2)$	-0.1313	0.0497	0.0085

Figure 6, both bar plots indicate that the optimal choice for u_3 is 4 while the bar plot of $\widehat{RC}_{3:3}(u_3)$ indicates that we might need to consider 1 and -5 as well. Therefore, we checked for u_4 and Figure 7 shows that no bar stands out in the graph of $\widehat{RC}_{4:4}(u_4)$ while the bar of $u_4 = 1$ is prominent in the graph $\widehat{IS}_{4:4}(u_4)$. Therefore, we took $X_1(t+1)X_2(t)$ as the fourth input.

We selected the lag terms as in Section 3 and 4.1 and the result of stepwise regression based on AIC is shown in Table 4. Four significant ($\alpha = 0.05$) interation terms, $x_1(t)x_2(t-4)$, $x_1(t-1)x_2(t-5)$, $x_1(t)x_2(t-1)$ and $x_1(t-1)x_2(t-2)$, are detected where the first two are from $X_1(t+4)X_2(t)$ and the last two are from $X_1(t+1)X_2(t)$.

5. Conclusion

Residual coherence and integrated spectrum proposed in this paper are graphical devices which point to possible significant interactions based on the result of Sections 3 and 4. Significant interactions could produce one or more than one prominent bars in the bar plots of $RC_{k:k}(u_k)$ and $IS_{k:k}(u_k)$ as functions of the kth input interaction.

When there are multiple prominent bars, one could consider u_{k+1} for more possible significant interactions. Once the input processes are determined, one can employ the regression-based selection method proposed in Khan, Katzoff and Kedem (2014) and Kedem (2016) to search for significant covariate interactions.

In addition, it is observed from the analysis in Section 4 that the cross product interaction $X_1(t+4)X_2(t)$ of the first order differences of OVX and GVX has significant influence on the first order differences of VIX and RVX. This suggests that daily increments of implied volatility of the stock market are possibly influenced by products of the daily increments (and their lags) of implied volatility of commodity markets. The process $X_1(t+4)X_2(t)$ might be an essential factor in the relationship between the implied volatility of stock market and certain commodity markets and therefore further exploration is warranted.

Acknowledgements

Research supported by a Faculty-Student Research Award, University of Maryland, College Park.

References

Brillinger, D. R., (1965). An introduction to polyspectra. *The Annals of mathematical statistics*, pp. 1351–1374.

Brillinger, D. R., Rosenblatt, M., (1967). Asymptotic theory of estimates of kth-order spectra. *Proceedings of the National Academy of Sciences of the United States of America*, 57(2), p. 206.

Elgar, S., Vanhoff, B., Aguirrre, L. A., Freitas, U. S., Chandran, V., (1998). Higher-order spectra of nonlinear polynomial models for chua's circuit. *International Journal of Bifurcation and Chaos*, 8(12), pp. 2425–2431.

Euan, C., Sun, Y., Ombao, H., (2019). Coherence-based time series clustering for statistical inference and visualization of brain connectivity. *Annals of Applied Statistics*, 13(2), pp. 990–1015.

Hinich, M. J., (1979). Estimating the lag structure of a nonlinear time series model. *Journal of the American Statistical Association*, 74 (366a), pp. 449–452.

Kedem, B., (2016). Coherence consideration in binary time series analysis. *Handbook of Discrete-Valued Time Series*, p. 311.

Kedem, B., Fokianos, K., (2002). *Regression models for time series analysis*. Wiley, Hoboken.

Kedem-Kimelfeld, B., (1975). Estimating the lags of lag processes. *Journal of the American Statistical Association*, 70(351a), pp. 603–605.

Khan, D., Katzoff, M., Kedem, B., (2014). Coherence structure and its application in mortality forecasting. *Journal of Statistical Theory and Practice*, 8(4), pp. 578–590.

Kimelfeld, B., (1974). Estimating the kernels of nonlinear orthogonal polynomial functionals. *The Annals of Statistics*, pp. 353–358.

Koopmans, L. H., (1974). *The spectral analysis of time series*. New Yorkn NY: Academic Press.

Maharaj, E. A., D'Urso, P., (2010). A coherence-based approach for the pattern recognition of time series. *Physica A: Statistical mechanics and its Applications*, 389(17), pp. 3516–3537.

Nikias, C. L., Mendel, J. M., (1993). Signal processing with higher-order spectra. *IEEE Signal processing magazine*, 10(3), pp. 10–37.

Sanaullah, M., (2013). A review of higher order statistics and spectra in communication systems. *Global Journal of Science Frontier Research, Physics and Space Science*, 13(4).

Sun, F. T., Miller, L. M., D'Esposito, M., (2004). Measuring interregional functional connectivity using coherence and partial coherence analyses of fmri data. *Neuroimage*, 21(2), pp. 647–658.

Tick, L. J., (1961). The estimation of "transfer functions" of quadratic systems. *Technometrics*, 3(4), pp. 563–567.

Whaley, R. E., (1993). Derivatives on market volatility: Hedging tools long overdue. *The journal of Derivatives*, 1(1), pp. 71–84.

Wiener, N., (1958). *Nonlinear problems in random theory*. M.I.T. Press, Cambridge.

APPENDIX

Since all G_k's are mutually orthogonal, fix k, then $\forall h$,

$$\begin{aligned}
&EG_k(t+h;u_1,\ldots,u_k)\overline{G_j(t;u_1,\ldots,u_j)}=0 \qquad j=1,\ldots,k-1\\
\Rightarrow &\int_{-\pi}^{\pi} e^{ih\lambda}\sum_{l=1}^{k} A_{l,k+l-1}(\lambda) f_{X_{l,u_l}X_{j,u_j}}(\lambda)d\lambda=0 \qquad j=1,\ldots,k-1\\
\Rightarrow &\sum_{l=1}^{k} A_{l,k+l-1}(\lambda) f_{X_{l,u_l}X_{j,u_j}}(\lambda)d\lambda=0 \qquad j=1,\ldots,k-1\\
\Rightarrow &\boldsymbol{F}_k(\lambda)\boldsymbol{A}_k(\lambda)=\boldsymbol{f}_k(\lambda)A_{k,1}(\lambda)
\end{aligned}$$

where $\boldsymbol{A}_k(\lambda)\equiv[A_{1,k}(\lambda),A_{2,k-1},\ldots,A_{k-1,2}]^T$. Then, by Cramer's rule,

$$A_{j,k-j+1}(\lambda)=\frac{\boldsymbol{F}_{k,j}(\lambda)}{\boldsymbol{F}_k(\lambda)}A_{k,1}(\lambda)=c_{k,j}(\lambda)A_{k,1}(\lambda)$$

for $j=1,\ldots,k-1$. Based on the orthogonality and the uniqueness of Fourier transform, we also have

$$\begin{aligned}
EG_k(t+h;u_1,\ldots,u_k)\overline{Y(t)}&=EG_k(t+h;u_1,\ldots,u_k)\overline{G_k(t;u_1,\ldots,u_k)}\\
\sum_{j=1}^{k} A_{k,k-j+1}(\lambda)f_{X_{j,u_j}Y}(\lambda)&=\overline{A_{k,1}(\lambda)}\sum_{j=1}^{k} A_{k,k-j+1}(\lambda)f_{X_{j,u_j}X_{k,u_k}}(\lambda)\\
A_{k,1}(\lambda)\sum_{j=1}^{k} c_{k,j}(\lambda)f_{X_{j,u_j}Y}(\lambda)&=|A_{k,1}(\lambda)|^2\sum_{j=1}^{k} c_{k,j}(\lambda)f_{X_{j,u_j}X_{k,u_k}}(\lambda)\\
A_{k,1}(\lambda)&=\overline{\left[\frac{\sum_{j=1}^{k} c_{k,j}(\lambda)f_{X_{j,u_j}Y}(\lambda)}{\sum_{j=1}^{k} c_{k,j}(\lambda)f_{X_{j,u_j}X_{k,u_k}}(\lambda)}\right]}
\end{aligned}$$

Therefore, all A's for G_k are solved.