Software Impacts 16 (2023) 100495

Contents lists available at ScienceDirect

Software Impacts

journal homepage: www.journals.elsevier.com/software-impacts

Original software publication

W2CWM2C reloaded: Ten years later

Josué M. Polanco-Martínez *

GECOS-IME, University of Salamanca, Salamanca, Spain
Basque Centre for Climate Change (BC3), Leioa, Spain

ARTICLE INFO

Keywords:
Wavelet correlation
Wavelet cross-correlation
Wavelet multiple correlation
Wavelet multiple cross-correlation
MODWT

ABSTRACT

We present an updated version v2.2 of the R package W2CWM2C (a graphical tool for wavelet (cross) correlation and multiple wavelet (cross) correlation analysis) v1.0. W2CWM2C contains a set of R functions that improve the graphical representations of wavelet correlation and cross-correlation for the bivariate and multivariate cases. The four functions contained in W2CWM2C are highly flexible since these functions contain several parameters to personalize the correlation heat maps and the format of these graphics, which can be displayed on the screen or can be saved as PNG, JPG, PDF or EPS. The W2CWM2C package also helps to handle the input data (multivariate time series) easily as a list of N columns and provides a multivariate dataset to exemplify its use. W2CWM2C is used for the first time to analyse multivariate climate time series..

Code metadata

Current code version	*v2.2*
Permanent link to code/repository used for this code version	https://github.com/SoftwareImpacts/SIMPAC-2023-14
Permanent link to Reproducible Capsule	https://codeocean.com/capsule/4105900/tree/v1
Legal Code License	*GPL (>= 2) License*
Code versioning system used	*None*
Software code languages, tools, and services used	*R*
Compilation requirements, operating environments & dependencies	*R (≥ 3.6), wavemulcor, waveslim, colorspace*
If available Link to developer documentation/manual	https://github.com/jomopo/W2CWM2C
Support email for questions	josue.m.polanco@gmail.com

1. Introduction to W2CWM2C

W2CWM2C is an R package of graphical tools that helps with the interpretation of the wavelet correlation and cross-correlation analysis for bivariate and multivariate time series [1,2]. The four R functions contained in *W2CWM2C* are based on the methods developed by [3,4]. The R package *W2CWM2C* V1.0 was published on CRAN (The Comprehensive R Archive Network) in November 2012 and its corresponding paper in 2014 [2], as well as the last update V.2.2 in January 2021 [1]. The main changes in the version V2.2 are related with some improvements in the heat maps and the format of these plot outputs (e.g., now it is possible to save these plots in four formats: PNG, JPG, PDF and EPS; plus via screen), improvements in the documentation, and last but not least, to fix some relevant bugs found by some users (look at the Acknowledgements).

This papers focus on the impact of the software *W2CWM2C* during the last ten years. We have decided to write a software update since *W2CWM2C* has been downloaded from CRAN around 33,000 times and the number of citations of its corresponding paper is close to 30 during the last ten years. These statistics do not seem very impressive, but downloads of *W2CWM2C* have remained stable over the first five years of its existence and have incremented conspicuously for the last five years. For these reasons, we think that our software still has a lot to offer to the scientific community. As the first application of *W2CWM2C* was in financial research and despite the wavelet correlation functions contained in our R package can be used with any type of time series that are not stationary, this has not been yet used in other scientific fields outside of economy, e.g., climate, ecology, geophysics, etc.

The functions contained in *W2CWM2C* are highly flexible since these contains some parameters to personalize the visualization of

The code (and data) in this article has been certified as Reproducible by Code Ocean: (https://codeocean.com/). More information on the Reproducibility Badge Initiative is available at https://www.elsevier.com/physical-sciences-and-engineering/computer-science/journals.

* Correspondence to: GECOS-IME, University of Salamanca, Salamanca, Spain.
E-mail addresses: josue.polanco@usal.es, josue.m.polanco@gmail.com.

https://doi.org/10.1016/j.simpa.2023.100495
Received 6 March 2023; Accepted 17 March 2023

the correlation heat maps. Furthermore, *W2CWM2C* also includes one dataset, named `dataexample` that contains seven European stock market indices, to exemplify its use. However, in this paper, the *W2CWM2C* package is used for the first time to analyse multivariate climate time series, and the results are shown in the illustrative example. This is done to introduce an application of *W2CWM2C* in other different scientific field outside of the economy and financial sciences. This will open up new opportunities to use the R package *W2CWM2C* to study other types of complex phenomena.

2. Software description with illustrative examples

The package *W2CWM2C* is written in R and does not require compilation or others external computational languages, and can be installed and used in the main operating systems (Windows, GNU/Linux and MacOS), since R is multiplatform. However, the *W2CWM2C* depends on the following R packages: *waveslim* [5], *wavemulcor* [6], and *colorspace* [7], all of which are available on CRAN. The first one is used to decompose the time series under analysis via the MODWT (Maximal Overlap Discrete Wavelet Transform) and to estimate the wavelet correlation for the bivariate case. The second package helps to estimate the wavelet correlation for the multivariate case, and the third one is used to build the palettes of the heat maps for the wavelet correlations. Moreover, *W2CWM2C* contains a relative moderate number of lines of code (ca. 450) in R, which were written specifically for this purpose. *W2CWM2C* is free software (GPL) and is freely available from CRAN [1], the official repository of R packages of the R project.

2.1. Software functionalities

The package *W2CWM2C* includes four functions: (1) the WC ("wavelet correlation") that estimates and plots as a heat map the wavelet correlation coefficients for the bivariate case for a group of time series (six as a maximum; for more than six series we suggest using the WMC function). That is, WC generates a heat map of the $C_{M,2}$ pairwise comparisons, which are plotted in descending order with respect to the sum of all the wavelet correlation coefficients for each pairwise comparison [2]. This function is based on the works of [4,5], in particular the *waveslim* `wave.multiple.correlation` function. (2) The WCC ("wavelet cross-correlation") function estimates and plots as a heat map the wavelet cross-correlation coefficients for two time series. This function is based on the work of [4,5], especially on the *waveslim* `spin.correlation` function. (3) The WMC ("wavelet multiple correlation") is used to estimate and to plot as a heat map the wavelet multiple correlation coefficients for the multivariate case. (4) The WMCC ("wavelet multiple cross-correlation") function estimates and plots as a heat map the wavelet multiple cross-correlation coefficients [3,6] for the multivariate case. The last two functions use the `wave.multiple.correlation` and `wave.multiple.cross.correlation` functions from the R package *wavemulcor* [3,6] to compute the wavelet multiple correlation and the wavelet cross-correlation for the multivariate case. The correlation coefficients that are not statistically significant (outside of the 95% CI) are plotted as blanks. All the functions use the `diverge_hcl` function from the R package *colorspace* to produce the palette of colours to be used in the correlation heat maps. A description of their syntaxes is as follows:

```
(1) WC(inputDATA, Wname, J, device="screen",
filename, Hgif, WFig, Hpdf, Wpdf)
```

- `inputDATA`: a numeric array containing M regular (evenly spaced) time series as a `ts` object.
- `Wname`: name of the wavelet filter used in the MODWT decomposition of the time series under study. There are several wavelet filters to be used, for example, it is common to use the Daubechies orthonormal compactly supported wavelet of length L = 8, i.e., "la8" [8].
- J: it is the maximum level of the MODWT decomposition and this must be an integer number. It is recommended to use `round` (log_2(N)), where N is the number of rows or elements of `inputDATA` [8,9].
- `device`: the type of output device for the plot output, by default is "screen", but it is also possible to use "jpg", "png", "pdf" and "eps".
- `filename`: the output filename to save heat maps.
- `Hfig/Wfig`: the height/width of the JPG or PNG image to save the wavelet correlation heat map.
- `Hpdf/Wpdf`: the height/width of the PDF or EPS format to save the wavelet correlation heat map.

The output of WC is a heat map of the wavelet correlation coefficients that are statistically significant (within the 95% confidence interval) and a numerical list that contains two elements:

- `wavcor.modwtsDAT`: it is an array of dimensions $C_{M,2}$ × J × 3, where the first element contains the wavelet correlation coefficients and the other two elements are the lower and upper bounds of the 95% confidence interval (CI) for these coefficients.
- `to3DpL`: this array also contains the wavelet correlation coefficients as `wavcor.modwtsDAT`, but their elements have been sorted in ascending order with respect to the wavelet scale J.

```
(2) WCC(inputDATA, Wname, J, lmax, device="screen",
filename, Hgif, WFig, Hpdf, Wpdf)
```

All of which have already been described except `lmax`, which is the maximum lag for which the wavelet cross-correlation is computed. By default is 30, but this value depends on the number of elements of the time series under study.

The output of WCC is a heat map for the wavelet cross-correlation coefficients that are statistically significant (within the 95% confidence interval) and a list that contains these coefficients for each lag, whose dimensions are given by (2 ·lmax + 1) × J.

```
(3) WMC(Wname, J, device="screen", filename, Hgif,
WFig, Hpdf, Wpdf)
```

All of which of these parameters have already been described. The output of the function WMC produces a heat map of the wavelet multiple correlation coefficients and a list (named LS), which contains two objects: `xy.mulcor` and YmaxR. The first one contains the correlation coefficients with their corresponding lower and upper bounds of the 95% confidence interval. The second object contains the index number of the variable/s that gives the maximum correlation against a linear combination of the rest of the variables under study [2,3].

```
(3) WMCC(Wname, J, lmax, device="screen", filename,
Hgif, WFig, Hpdf, Wpdf)
```

The parameters in WMCC have already been defined. The graphical output is a heat map that shows the wavelet correlation coefficients statistically significant (95% confidence), which includes small vertical dashed lines to indicate the maximum correlation value for each J level. This plot also includes a label to indicate the variable/s that maximize the cross-correlation against a linear combination of the rest of the variables for each J level [2,3]. On the other hand, the numerical output is a list of two elements as in WMC but this list also includes information about the lags.

Finally, the dataset included in *W2CWM2C*, named `dataexample`, must be loaded manually (i.e. `>data(dataexample)`). `dataexample` contains seven European stock market indices (daily closing prices) spanning from January 2, 2004 to June 29, 2012 and come from Yahoo finance [1,2].

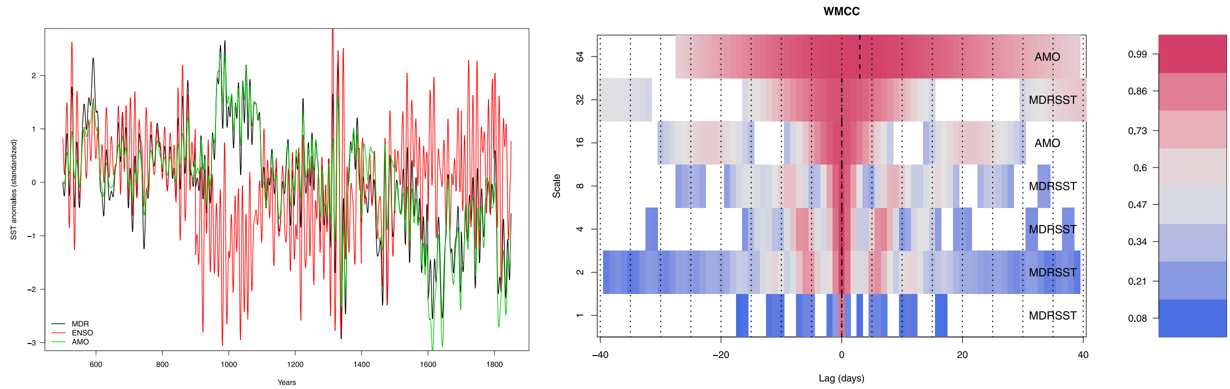

Fig. 1. Climate time series (left) and the heat map (right) of the wavelet multiple cross-correlation coefficients obtained through the function WLMC from the R package *W2CWM2C* [1]. Wavelet correlation coefficients that are not statistically significant (outside of the 95% CI) are in blank. The long dashed vertical lines indicate where in time the highest wavelet correlation values are localized. The labels in the corner right side are the climate variables that maximize the wavelet multiple cross-correlation for each wavelet scale.

```
load("data_climate")
inputDATA <- ts(data_climate[,2:4], start=1,
   frequency=1)
Wname      <- "la8"
J          <- 6
lmax       <- 40
wmcc       <- WMCC(inputDATA, Wname=Wname, J=J,
                lmax=lmax, device="pdf",
                filename="example_WMCC",
                Wpdf=12, Hpdf=9)
```

Listing 1: Code used to produce Fig. 1.

2.2. Illustrative example

The illustrative example (Fig. 1) shows the output of the function WLMC. The R code used to produce this figure is in the Listing 1. The climate variables are in anomalies of temperature (°C), cover the years from 500 to 1850 and are paleoclimate reconstructions of MDRSST (Sea Surface Temperatures from the Main Developed Region — DMR for tropical cyclones) [10], ENSO (El Niño-Southern Oscillation SST, el Niño 3 region) [11], and AMO (The North Atlantic Multidecadal Oscillation SST) [11]. These climate variables show a significant level of correlation around the lag-0 for all the wavelet scales such as it is expected since the three climate variables are closely related at different time-scales [8,10–12]. Furthermore, the main dominant climate variable is the MDRSST and to lesser extent the AMO. This result is in concordance with a recent study published by [8], though they used a dynamic multiple wavelet correlation approach. For a deeper scientific discussion about this climate data analysis look at the aforementioned publication.

3. Impact

The computational package *W2CWM2C* is intended for scientists and researchers who would like to conduct applied research on time series analysis, especially to find relationships between two or more variables. *W2CWM2C* is easy to use, is well documented and is available freely from CRAN [1]. The first application of *W2CWM2C* was to analyse financial time series [2], perhaps for this reason this software has been mainly used within this field despite this can be used in other scientific disciplines. Till the date, *W2CWM2C* has been downloaded 33192 (on average 3300 times per year) times from CRAN (Fig. 2) (this statistic does not take into account other mirrors or repositories different from CRAN). Fig. 2 (left) reveals that the number of downloads has remained relatively stable for the first five years with a moderate trend to increase for the last five years. However, total yearly sum of downloads (Fig. 2 right) increased conspicuously between 2019 and 2021 reaching its maximum value in 2020 with 6652 downloads, though this number has decreased to 3385 in 2022. On the other hand, total annual citations from Google Scholar (look at Refs. [13–23,23–39] for the citations) for the time interval 2012—2022 is 29 (Fig. 3), with a mean value of approximately three citations per year. Fig. 3 also reveals that the number of publications has increased considerably since 2018 with a slight trend to decrease in 2022, but slightly above of the average value of the citations. There seems to be an apparent relationship between the number of downloads and citations per year, i.e., an increases in the number of downloads perhaps implies an increases in the number of citations. Although the opposite might also be true, that is, an increases in the number of published papers and citations can also provoke an increases in the number of downloads and use of *W2CWM2C*. Unfortunately to address this hypothesis is beyond the aim of this paper and it is not possible to probe this quantitatively due to the limited number of data. Nevertheless from a pragmatic point of view, the evidence of the impact of the software *W2CWM2C* in terms of downloads, citations and use is quite clear. This indicates that the R package *W2CWM2C* has an enormous potential not only to continue being used in financial research, but also start being used in others scientific and engineering fields, such as climatology and related areas of research.

4. Conclusion and future development

Ten years after the publication of the R package *W2CWM2C*, this has been downloaded 33192 times from CRAN and has been cited 29 times from Google Scholar. Both, the number of downloads and citations has increased during the last 5 years indicating that *W2CWM2C* is still alive. The main differences between the last version of *W2CWM2C* with the first version is related with improvements in the correlation heat maps and the format of these plots, in the documentation, and to fix some relevant bugs found by some users. Furthermore, this software paper introduces for the first time the *W2CWM2C* to analyse multivariate climate time series, opening up new opportunities to use this package in other fields outside of economy and financial sciences.

Our future plan includes implementation of more functions to be included in *W2CWM2C*. For instance, one to address irregular (unevenly spaced) time series that are very common in diverse branches of sciences and engineering (e.g. ecology, paleoclimate, geology or oceanography). However, from our particular point of view, the most imperative future direction of research will include in *W2CWM2C* a causality test to infer causation between time series. This will be helpful

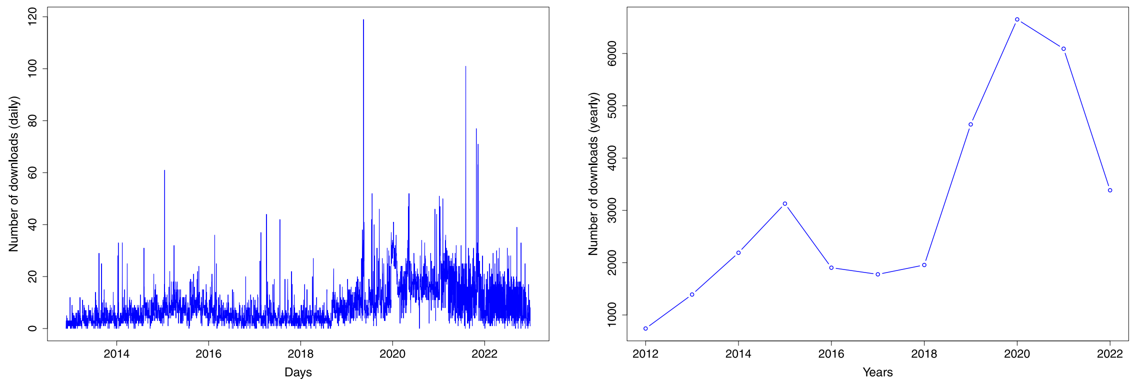

Fig. 2. Number of downloads from CRAN for the R package *W2CWM2C* with daily (left) and yearly (cumulative sum; right) data spanning the time interval 24/11/2012—31/12/2022.

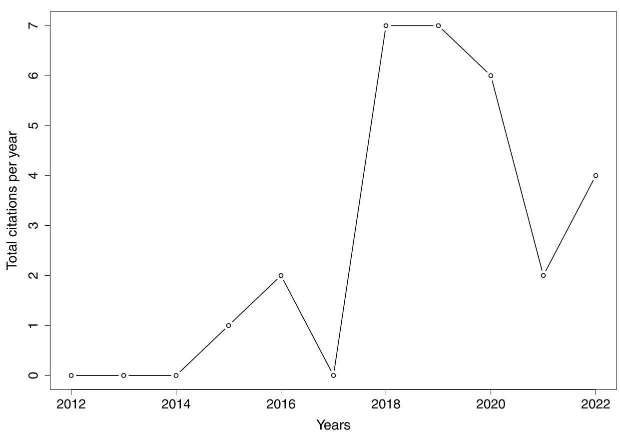

Fig. 3. Number of citations per year for the R package *W2CWM2C* spanning the time interval 24/11/2012—31/12/2022. Data come from Google Scholar.

to find out the direction of causation when two or more time series are analysed and it is not so evident the cause–effect relationship between these series (variables).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

JMPM acknowledges to the Excellence Unit GECOS, Spain (reference number CLU-2019-03), Universidad de Salamanca for funding support. I am thanks to Debojyoti Das, John Garrigan, Peterson Owusu Junior, Rim Ibrahim, and Ato Wilberforce for reporting some bugs found in W2CWM2C package.

References

[1] J.M. Polanco-Martinez, W2CWM2C: A graphical tool for wavelet (cross) correlation and wavelet multiple (cross) correlation analysis, 2021, URL https://CRAN.R-project.org/package=W2CWM2C. R package version 2.2.
[2] J.M. Polanco-Martínez, F.J. Fernández-Macho, Package W2CWM2C: description, features, and applications, Comput. Sci. Eng. 16 (6) (2014) 68–78.
[3] J. Fernández-Macho, Wavelet multiple correlation and cross-correlation: A multiscale analysis of Eurozone stock markets, Phys. A 391 (4) (2012) 1097–1104.
[4] B. Whitcher, P. Guttorp, D.B. Percival, Wavelet analysis of covariance with application to atmospheric time series, J. Geophys. Res.: Atmos. 105 (D11) (2000) 14941–14962.
[5] B. Whitcher, The waveslim package, 2022, URL https://CRAN.R-project.org/package=waveslim. R package version 1.8.4.
[6] J. Fernández-Macho, Wavemulcor: Wavelet routines for global and local multiple correlation, 2021, URL https://CRAN.R-project.org/package=wavemulcor. R package version 3.1.2.
[7] R. Ihaka, R. Murrell, K. Hornik, J. Fisher, R. Stauffer, W. CO, C. McWhite, A. Zeileis, Colorspace: A toolbox for manipulating and assessing colors and palettes, 2022, URL https://CRAN.R-project.org/package=colorspace. R package version 2.0.3.
[8] J. Polanco-Martínez, J. Fernández-Macho, M. Medina-Elizalde, Dynamic wavelet correlation analysis for multivariate climate time series, Sci. Rep. 10 (1) (2020) 1–11.
[9] R. Gençay, F. Selçuk, B.J. Whitcher, An introduction to wavelets and other filtering methods in finance and economics, Academic Press, 2001.
[10] M. Mann, J. Woodruff, J.P. Donnelly, Z. Zhang, Atlantic hurricanes and climate over the past 1,500 years, Nature 460 (7257) (2009) 880–883.
[11] M. Mann, Z. Zhang, S. Rutherford, R. Bradley, M. Hughes, D. Shindell, C. Ammann, G. Faluvegi, F. Ni, Global signatures and dynamical origins of the little ice age and medieval climate anomaly, Science 326 (5957) (2009) 1256–1260.
[12] M. Medina-Elizalde, J.M. Polanco-Martínez, F. Lases-Hernández, R. Bradley, S. Burns, Testing the “tropical storm” hypothesis of Yucatan Peninsula climate variability during the Maya Terminal Classic Period, Quat. Res. 86 (2) (2016) 111–119.
[13] J. Polanco-Martínez, S.H. Faria, Towards a new statistical tool for analyzing unevenly spaced paleoclimate time series, BC3 Working Paper Series 2014-07, Bilbao, Spain, 2015.
[14] W.-Y. Chen, Y.-H. Lin, Co-movement of healthcare financing in OECD countries: Evidence from discrete wavelet analyses, Romanian J. Econ. Forecast. 19 (3) (2016) 40–56.
[15] J.M. Polanco-Martínez, L.M. Abadie, Analyzing crude oil spot price dynamics versus long term future prices: A wavelet analysis approach, Energies 9 (12) (2016) 1089.
[16] J.M. Polanco-Martínez, J. Fernández-Macho, M. Neumann, S.H. Faria, A pre-crisis vs. crisis analysis of peripheral EU stock markets by means of wavelet transform and a nonlinear causality test, Phys. A 490 (2018) 1211–1227.
[17] J. Polanco-Martínez, L.M. Abadie, J. Fernández-Macho, A multi-resolution and multivariate analysis of the dynamic relationships between crude oil and petroleum-product prices, Appl. Energy 228 (2018) 1550–1560.
[18] G. Tweneboah, P. Alagidede, Interdependence structure of precious metal prices: A multi-scale perspective, Resour. Policy 59 (2018) 427–434.
[19] J.M. Polanco-Martínez, S. Faria, Estimation of the significance of the Foster's wavelet spectrum by means of a permutation test and its application for paleoclimate records, Boletín Geol. y Minero 129 (3) (2018) 549–564.
[20] D. Das, M. Kannadhasan, K.H. Al-Yahyaee, S.-M. Yoon, A wavelet analysis of co-movements in Asian gold markets, Phys. A 492 (2018) 192–206.
[21] D. Das, P. Bhowmik, R. Jana, A multiscale analysis of stock return co-movements and spillovers: Evidence from Pacific developed markets, Phys. A 502 (2018) 379–393.
[22] J. Fernandez-Macho, Package wavemulcor: Wavelet multiple regression and correlation in R, Comput. Sci. Eng. 21 (6) (2018) 63–73.
[23] M.U. Rehman, Dynamic correlation pattern amongst alternative energy market for diversification opportunities, J. Econ. Struct. 9 (1) (2020) 1–24.
[24] R. Nagayev, Y. Dinç, M. Çakır, Multiresolution analysis of comovements between foreign exchange and equity markets: evidence from Turkey, in: 15th International Conference on Knowledge, Economy & Management, Bilal Kargı, 2019.
[25] A.K. Tiwari, M. Shahbaz, H.M. Hasim, M.M. Elheddad, Analysing the spillover of inflation in selected Euro-area countries, J. Quant. Econ. 17 (3) (2019) 551–577.
[26] G. Tweneboah, Dynamic interdependence of industrial metal price returns: evidence from wavelet multiple correlations, Phys. A 527 (2019) 121153.

[27] D. Das, K. Manoharan, Emerging stock market co-movements in South Asia: wavelet approach, Int. J. Manag. Finance 15 (2) (2019) 236–256.
[28] J.M. Polanco-Martínez, Dynamic relationship analysis between NAFTA stock markets using nonlinear, nonparametric, non-stationary methods, Nonlinear Dynam. 97 (1) (2019) 369–389.
[29] K. Ijasan, G. Tweneboah, M. Omane-Adjepong, P. Owusu, On the global integration of REITs market returns: A multiresolution analysis, Cogent Econ. Finance 7 (1) (2019) 1690211.
[30] M. Omane-Adjepong, I.P. Alagidede, Dynamic linkages and economic role of leading cryptocurrencies in an emerging market, Asia-Pacific Financial Mark. 27 (4) (2020) 537–585.
[31] M.U. Rehman, X.-V. Vo, Is a portfolio of socially responsible firms profitable for investors? J. Sustain. Finance Invest. 10 (2) (2020) 191–212.
[32] A. Bhandari, B. Kamaiah, Time-varying nature of stock market interdependence: A global perspective, Econ. Political Wkly. 55 (2020).
[33] M.U. Rehman, X.V. Vo, Do alternative energy markets provide optimal alternative investment opportunities? N. Am. J. Econ. Finance 54 (2020) 101271.
[34] B. Kaffel, F. Abid, Time-frequency wavelet analysis of stock-market co-movement between and within geographic trading blocs, in: Handbook of Financial Econometrics, Mathematics, Statistics, and Machine Learning, World Scientific, 2021, pp. 1399–1437.
[35] Y. Zhu, Essays on Agricultural Commodity Prices (Ph.D. thesis), Newcastle University, UK, 2021.
[36] T. Saeed, N. Nautiyal, M.U. Rehman, H.H. Alsulami, M.S. Alhothuali, Do gulf stock markets share time varying connectedness, Appl. Econ. (2022) 1–19.
[37] M.U. Rehman, N. Nautiyal, Dependence and spillover between islamic stock markets under different market conditions, 2022, Preprint available at SSRN.
[38] S. Coronado, J.N. Martinez, V. Gualajara, O. Rojas, Transfer entropy granger causality between news indices and stock markets in US and Latin America during the COVID-19 pandemic, Entropy 24 (10) (2022) 1420.
[39] A.N.K. Adjei, G. Tweneboah, P. Owusu, Interdependence of economic policy uncertainty and business cycles in selected emerging market economies, J. Financial Econ. Policy 14 (5) (2022) 601–632.