

# Financial Time Series Clustering

Kartikay Gupta<sup>(✉)</sup> and Niladri Chatterjee

Indian Institute of Technology Delhi, New Delhi, India  
{maz158144,niladri}@iitd.ac.in

**Abstract.** Financial time series clustering finds application in forecasting, noise reduction and enhanced index tracking. The central theme in all the available clustering algorithms is the dissimilarity measure employed by the algorithm. The dissimilarity measures, applicable in financial domain, as used or suggested in past researches, are correlation based dissimilarity measure, temporal correlation based dissimilarity measure and dynamic time wrapping (DTW) based dissimilarity measure. One shortcoming of these dissimilarity measures is that they do not take into account the lead or lag existing between the returns of different stocks which changes with time. Mostly, such stocks with high value of correlation at some lead or lag belong to the same cluster (or sector). The present paper, proposes two new dissimilarity measures which show superior clustering results as compared to past measures when compared over 3 data sets comprising of 526 companies. *abstract* environment.

**Keywords:** Clustering · Finance · Time series · Cross-correlation · Forecasting

## 1 Literature Review

In the last two decades a significant amount of work has been done on time series clustering [7]. Financial Time series clustering is a subject that has also gained lots of attention in the last decade [14]. It is an important area of research that finds wide applications in noise reduction, forecasting and enhanced index tracking [1]. A good forecast of future prices is desired by financial companies especially the algorithmic trading firms. Index track funds are low cost funds that closely track the returns of one particular index. In a typical financial time series clustering procedure, each time series is considered as an individual object and inter-object dissimilarities are then calculated. Subsequent clustering is done using one of the many clustering algorithms. Rosario [2] used a well-known dissimilarity measure based on the Pearson's correlation coefficient. They then cluster the data set using single linkage hierarchical clustering. Saeed [4] used symbolic representation for dimensionality reduction of financial time series data, and then used longest common subsequence as similarity measurement. Guan [5] proposed a similarity measure for time series clustering which relied on the signs (positive or negative) of the logarithmic returns of the stock

prices. It did not take into account the size of movement. Marti et al. [6] tried to find answer to the question, what should be an appropriate length of a time series for the clustering procedure. John et al. [8] proposed a shape based dissimilarity for time series clustering which relied on the cross correlation coefficients. This work comes closest to our approach, but still there is a significant difference between the two approaches. While calculating their dissimilarity measure, they do not break the time series into smaller segments. In the present approach, time series are broken into smaller parts and then a different procedure is followed. This is because lead – lag relationship between two time series may change over time. Further, we propose another dissimilarity measure that gives similar or better results as compared to the first dissimilarity measure proposed in the present paper.

## 2 Preliminaries

In the present work, hierarchical clustering is used to form clusters from the inter object dissimilarity matrix computed using a dissimilarity measure. Linkage method is an important aspect of hierarchical clustering algorithm. In the present work, we choose ‘single’ linkage and ‘ward’ linkage (implemented as ward.D2 linkage in R ‘stats’ package) for our analysis. This is because ‘single’ linkage has been a preferred choice of researchers in financial time series clustering papers, e.g., [2, 3]. Ward linkage [10] was used for financial time series clustering by Guan [5]. Details about pre-existing dissimilarity measures i.e., Correlation based dissimilarity measure (COR) and Temporal correlation based dissimilarity measure (CORT) are given in the Appendix. Additionally, taking inspiration from [16], lead/lag time between two time series  $X_T$  and  $Y_T$  is defined as the integer ‘k’ which maximizes cross correlation between the two time series.

## 3 Proposed Dissimilarity Measures

### 3.1 Cross-Correlation Type Dissimilarity Measure (CCT)

Given two time series  $X_T$  and  $Y_T$  (each of length  $T$ ), here the interest is in finding a dissimilarity measure between them. Let ‘m’ be the maximum value of lead or lag being taken into consideration for calculation of dissimilarity between the two time series. We consider a segment of time series  $X_T$  starting from ‘m+1’ and ending at ‘T-m-1’, and divide this segment into ‘n’ equal parts each of length ‘p’. Here, we conveniently choose ‘m’, ‘n’ and ‘p’ such that  $2m+np+1 = T$ . This is required to make sure that all the data points in time series are utilized for the calculation of dissimilarity measure. Though it would still suffice if ‘2m+np+1’ is slightly less than  $T$ . Another couple of restrictions are that  $m \leq p$  and  $p \geq 15$ . These restrictions have been imposed to avoid unwanted cross-correlations. Now, with this background we define our CCT similarity measure which is given by the following:

$$CCT = \frac{1}{n} \times \sum_{l=1}^n \max\{CCT_k(m+1+p \times (l-1)) \mid -m \leq k \leq m\}, \quad (1)$$

where  $CCT_k(i)$  is defined as follows

$$CCT_k(i) = \frac{\sum_{j=i}^{i+p-1} (x_{j+1} - x_j)(y_{j+k+1} - y_{j+k})}{\sqrt{\sum_{j=i}^{i+p-1} (x_{j+1} - x_j)^2} \sqrt{\sum_{j=i}^{i+p-1} (y_{j+k+1} - y_{j+k})^2}} \tag{2}$$

The value of  $CCT_k$  is same as the correlation between returns of a segment of time series  $Y_t$  at a lead  $k$  with respect to a segment of time series  $X_t$ . The motivation behind this similarity measure is that similar financial time series would have more sub-segments which are highly correlated with each other at some lead or lag. Stock prices of similar assets in different exchanges exhibit such pattern. This phenomena is perfectly captured by CCT similarity measure. Notice that value of each of  $CCT_k(i)$  lies in the interval  $[-1,1]$ . Hence, the value of CCT similarity measure also lies in the interval  $[-1,1]$ . This similarity measure is then converted into a dissimilarity measure using a function, which is given by:

$$\phi(u) = \frac{2}{1 + e^{4 \times u}} \tag{3}$$

We choose to do our analysis with the above function as opposed to the function  $\sqrt{2(1-x)}$  for conversion of similarity measure into dissimilarity measure. Through our data experiments, we see that even this function gives similar results as the function in Eq. 3. Though we have not given details of those experiments in this paper. The dissimilarity measure thus obtained after conversion can be used for clustering of financial time series data.

As an example for CCT similarity measure consider the two time series data given in Table 1. Since the length of each time series is 37, one suitable choice for ‘p’, ‘n’ and ‘m’ may be 15, 2 and 3, respectively. In subsequent calculations it is found that for the first segment of series 1 (i.e., 4th to 18th data point in Series 1) the max  $CCT_k(4)$  value is 0.57 which exists at  $k = 3$ . For the second segment (i.e., 19th to 33rd data point in Series 1) the max  $CCT_k(19)$  value is 0.27 which exists at  $k = 0$ . Thus the value of CCT measure comes out to be 0.42.

**Table 1.** Hypothetical data set for two time series each of length 37.

S. no.	Series 1	Series 2	S. no.	Series 1	Series 2	S. no.	Series 1	Series 2	S.no.	Series 1	Series 2
1	0.81	0.15	11	0.96	0.96	21	0.58	0.68	31	0.13	0.47
2	0	0.76	12	0.69	0.36	22	0.2	0.83	32	0.23	0.14
3	0.92	0.01	13	0.05	0.7	23	0.69	0.01	33	0.75	0.98
4	0.94	0.1	14	0.85	0.66	24	0.76	0.01	34	0.63	0.85
5	0.97	0.56	15	0.55	0.37	25	0.77	0.44	35	0.29	0.36
6	0.77	0.88	16	0.65	0.07	26	0.13	0.01	36	0.33	0.55
7	0.26	0.91	17	0.39	0.63	27	0.89	0.05	37	0.65	0.39
8	0.68	0.05	18	0.32	0.35	28	0.78	0.61			
9	0.13	0.26	19	0.7	0.79	29	0.95	0.69			
10	0.82	0.5	20	0.39	0.13	30	0.56	0.13			

### 3.2 Cross-Correlation Type-II Dissimilarity Measure (CCT-II)

In another version of this dissimilarity measure, we ignore those intervals whose ‘maximum  $CCT_k(i)$ ’ value is less than a given threshold (denoted by ‘Thr’). We mean that instead of the ‘maximum  $CCT_k(i)$ ’ value we put ‘0’, when that is less than the threshold. Rest of the computations remains the same. The expression for this similarity measure is as follows:

$$\text{CCT-II} = \frac{1}{n} \times \sum_{l=1}^n (M(m+1+p \times (l-1)) \times I_{[Thr,1]}(M(m+1+p \times (l-1)))) \quad (4)$$

$$M(i) = \max\{CCT_k(i) \mid -m \leq k \leq m\} \quad (5)$$

where  $CCT_k(i)$  is same as defined in (3)

and  $I_{[Thr,1]}(x)$  denotes the indicator function of the set  $[Thr, 1]$ .

This similarity measure is then converted into dissimilarity measure using the function given in Eq. (3).

As an example, consider the time series data given in Table 1. We now compute CCT-II similarity measure with the same set of values (as used for computing CCT similarity measure) i.e.  $p = 15$ ,  $n = 2$  and  $m = 3$ . The threshold value is set as 0.50. Again it is noticed that  $\max CCT_k()$  values are 0.57 and 0.27. But since 0.27 lies below the threshold thus it is replaced by 0. Thus, the value of CCT-II similarity measure comes out to be 0.29.

The motivation behind this dissimilarity measure is that often the stock prices are weakly correlated to each other at some lead or lag. This may be due to a general trend followed by all stocks or it may be a spurious relation. This leads to error while clustering the data set using CCT dissimilarity measure. This noise can be removed by not considering the value of cross-correlations whose value is less than a given threshold.

In order to apply CCT-II measure of dissimilarity between two time series we need to fix the threshold value beforehand. The threshold value obviously lies in the interval  $(0, 1)$ . Additionally, it can’t be close to 0 as then there would remain very little difference between the values of CCT and CCT-II dissimilarity measures. It can’t be close to 1 as then most of the information regarding the behavior of two time series would be ignored and hence will not be considered in the evaluation of the final expression of CCT-II dissimilarity measure. Thus, threshold value should be kept close to 0.5. For this work we have chosen the range for the threshold to be  $[0.35, 0.65]$ . If a test data set is available then it can be used to find the optimal value of threshold. If test data set is not available, then choose threshold value as any random value in the range  $[0.35, 0.65]$ .

The time complexity of CCT and CCT-II dissimilarity measure is  $O(pmn)$ . Calculation of  $CCT_k(i)$  value requires  $O(p)$  computations, as we are calculating cross-correlation value at lead ‘k’ between two time series segments each of length ‘p’. Since we need to evaluate this expression for all k such that  $-m \leq k \leq m$ , thus it takes  $O(pm)$  time to evaluate ‘ $\max\{CCT_k(i) \mid -m \leq k \leq m\}$ ’. This

process is repeated ‘n’ times thus, CCT dissimilarity measure is of  $O(pmn)$  time complexity. Similarly, it can be argued that CCT-II dissimilarity measure is of  $O(pmn)$  time complexity. This time complexity is better than time complexity of Dynamic Time Wrapping dissimilarity measure, which takes  $O(T^2)$  time [15] to find its final expression, where T is the length of time series. Since ‘ $np+2m+1$ ’ is equal to T or slightly less than T, thus, this time complexity is equivalent to  $O(n^2p^2)$ . This time complexity is clearly greater than  $O(pmn)$  as  $m \leq p$ . Lower time complexity of proposed measures enable its computations to be carried out faster.

## 4 Experiments and Analysis

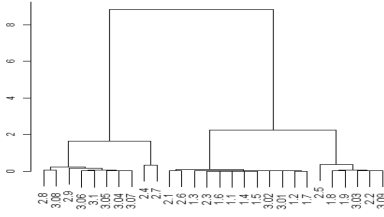
Experiments are conducted on 3 data sets one by one. Each data set consists of ‘End of Day’ (EOD) stock prices of some companies. Figure 1 depicts the format of data sets used for the experiments in the present paper. While clustering, each time series associated with a company is considered as an individual object. Inter object dissimilarities are then calculated with each of the 4 measures i.e., COR, CORT, CCT, CCT-II. A dissimilarity matrix is thus created, which is used for further analysis. Single linkage hierarchical clustering and Ward linkage hierarchical clustering are then employed to create the corresponding dendograms (see Fig. 2). The dendogram then can be cut at any level to form desired number of clusters of the data set.

Company Name	Date 1	Date 2	Date 3		Date T
Company 1	0.45	0.46	0.43		0.52
Company 2	1.56	1.59	1.58		1.42
Company 3	2.89	2.85	2.91	• • •	2.50
⋮					
Company N	1.07	1.08	1.09		1.05

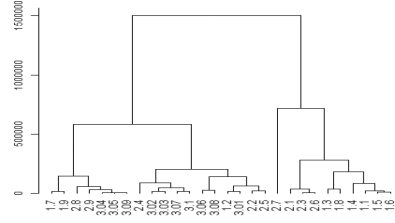
Fig. 1. EOD stock prices of the companies.

A cluster evaluation measure is then used to compare the clustering results obtained through the different dissimilarity measures. The present paper uses the cluster evaluation measure defined in [9, 16]. This cluster evaluation measure lies in the range [0,1]. Higher value of cluster evaluation measure corresponds to better clustering results.

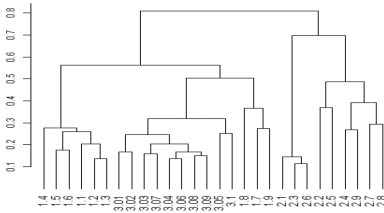
Here, we discuss results corresponding to each data set in a different subsection. R version 3.1.1 was used for preparation of all the figures presented in this paper. Since the number of data points in all time series in all the experiments is 2014, hence the parameter values of ‘n’, ‘p’ and ‘m’ are taken to be 18,100 and 100 respectively in all the experiments mentioned below. In these



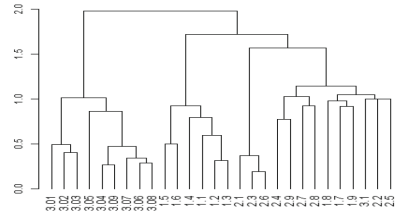
(a) Dissimilarity Measure: COR



(b) Dissimilarity Measure: CORT



(c) Dissimilarity Measure: CCT



(d) Dissimilarity Measure: CCT-II

**Fig. 2.** The above figures give dendrograms corresponding to different dissimilarity measures. Ward linkage hierarchical clustering has been used for forming the dendrograms. Each object in the clusters is represented by a number whose unit's place indicate it's true cluster value.

experiments, the threshold value for the CCT-II measure concerning Indian companies (the first data set) is taken to be higher as compared with the companies traded in USA (the second and third data set). This is because prices of the Indian companies tend to show more noise or spurious cross-correlations as compared to prices of American companies. Thus, threshold value for the first data set is taken to be 0.65 and threshold value for the second & third data set is taken to be 0.35.

### 4.1 Indian Data Set

The EOD stock prices of the companies listed in Table 2, form the first data set of our experiments. The time-span of the prices is from 5th August 2008 till 2nd May 2016. This data set can be originally clustered into 3 clusters as indicated in Table 2. Dendrograms obtained on this data set using Ward linkage hierarchical clustering are shown in Fig. 2. Table 3 gives cluster evaluation measure corresponding to different representations (number of clusters formed through dendrogram) and different dissimilarity measures. Amongst all the dendrograms shown in the Fig. 2, the best result is seen in Fig. 2 (d). Figure 2(d) shows dendrogram formed using CCT-II dissimilarity measure. Cluster evaluation measure

**Table 2.** The name of the companies whose stock prices are part of the Indian data set. These companies can be divided into 3 broad categories.

Cluster 1		Cluster 2		Cluster 3	
Top Private Banks in India		Top Oil Companies in India		Top Public Banks in India	
Name	Symbol	Name	Symbol	Name	Symbol
HDFC	1.1	Indian Oil Corp. Ltd.	2.1	SBI	3.01
ICICI	1.2	ONGC	2.2	Bank of Baroda	3.02
AXIS	1.3	Bharat Petroleum	2.3	PNB	3.03
Kotak Mahindra Bank Ltd.	1.4	Essar oil Ltd.	2.4	IDBI Bank Ltd.	3.04
Indusind Bank Ltd.	1.5	Cairn India Ltd.	2.5	Central Bank of India	3.05
Yes Bank Ltd.	1.6	Hindustan Petroleum Corp. Ltd.	2.6	Canara Bank	3.06
The Federal Bank Ltd.	1.7	Aban offshore Ltd.	2.7	Union Bank of India	3.07
Karur Vysya Bank Ltd.	1.8	Hindustan Oil exploraion company Ltd.	2.8	Bank Of India	3.08
South Indian Bank Ltd.	1.9	Mangalore Refinery & Petrochemicals Ltd.	2.9	Syndicate Bank	3.09
				Indian Bank	3.10

**Table 3.** Cluster evaluation measure for the different dissimilarity measures. This table corresponds to the Indian data set. ‘Number of Clusters’ represent the number of clusters formed using the dendogram. Higher the measure the closer is the representation to the true clusters. ‘COR’, ‘CORT’, ‘CCT’ and ‘CCT-II’ stand for correlation based dissimilarity measure, temporal correlation based dissimilarity measure, first proposed measure and second proposed measure respectively.

Cluster evaluation measure				
Number of clusters	COR	CORT	CCT	CCT-II
2	0.56	0.59	0.78	0.74
3	0.56	0.58	<b>0.71</b>	<b>0.86</b>
4	0.57	0.56	0.82	0.79

verifies that the clusters obtained through the proposed measures are better than the pre-existing measures. Cluster evaluation measure is consistently higher for the proposed measures, irrespective of how many clusters we form out of the dendograms.

Clustering results obtained on first data set using Single linkage hierarchical clustering also verify the superiority of CCT and CCT-II measures. In this case also, Cluster evaluation measure values for CCT and CCT-II measures are higher than the COR and CORT dissimilarity measures.

## 4.2 S&P500 Data Set

The EOD stock prices of the companies listed in S&P500 index form the second data set. These companies are traded in USA. The time span of the prices for this data set is from 12th September 2008 till 23rd august 2016. The companies

whose prices were not available for the complete time span under consideration, were removed from the data sets. This data set can be originally clustered into 10 clusters, where each cluster represents the sector of the company.

In the experiment associated with this data set, Cluster evaluation measure values are given in Tables 4 and 5. The threshold value is taken to be 0.35 for CCT-II measure in this experiment. In the case of ward linkage (as seen in Table 4) CCT and CCT-II measures clearly give better results as compared to COR and CORT dissimilarity measures. In the case of single linkage (as seen in Table 5), all measures give similar results.

**Table 4.** Cluster evaluation measure corresponding to different dissimilarity measures and number of clusters. Ward linkage hierarchical clustering has been used for clustering the datasets. This table corresponds to the S&P500 data set which can be originally divided into 10 clusters.

Cluster evaluation measure				
Number of clusters	COR	CORT	CCT	CCT-II
9	0.23	0.24	0.69	0.65
10	0.23	0.24	<b>0.69</b>	<b>0.70</b>
11	0.23	0.24	0.69	0.68

**Table 5.** Cluster evaluation measure corresponding to different dissimilarity measures and number of clusters. Single linkage hierarchical clustering has been used for clustering the datasets. This table corresponds to the S&P500 data set which can be originally divided into 10 clusters.

Cluster evaluation measure				
Number of clusters	COR	CORT	CCT	CCT-II
9	0.18	0.18	0.21	0.21
10	0.18	0.18	<b>0.21</b>	<b>0.21</b>
11	0.18	0.18	0.21	0.21

**Table 6.** Cluster evaluation measure for the different dissimilarity measures. ‘Number of Clusters’ represent the number of clusters formed using the dendrogram. Higher value corresponds to better clustering results. This table corresponds to the DJIA data set which can be originally divided into 8 clusters.

Cluster evaluation measure				
Number of clusters	COR	CORT	CCT	CCT-II
7	0.39	0.39	0.67	0.54
8	0.40	0.43	<b>0.65</b>	<b>0.56</b>
9	0.42	0.43	0.66	0.58



### 4.3 DJIA Data Set

The EOD stock prices of the companies listed in Dow Jones Industrial Average index form the third data set of our experiments. The time span of the prices for this data set is from 12th September 2008 till 23rd august 2016. This data can be originally clustered into 8 clusters. Here, the threshold value is taken to be 0.35 for CCT-II measure.

As can be seen in Table 6, the proposed measures show better or similar results as compared to pre-existing measures when hierarchical clustering is done using Ward linkage. In the case, when clustering is done using Single linkage hierarchical clustering for this data set, the proposed measures give similar or slightly inferior results as compared to COR & CORT measures. Though in this case also, maximum value of cluster evaluation measure (amongst all possible number of clusters) is seen corresponding to CCT measure.

Now, we discuss some of the directions in which future work could be carried out related to this research paper. Determination of optimal values of threshold,  $n$ ,  $p$  and  $m$  could be carried out in future. Also, optimal function for conversion of similarity measure into dissimilarity measure needs to be determined.

## 5 Conclusion

Financial time series clustering is an important area of research and finds wide applications in noise reduction, forecasting and index tracking. In this paper, two new dissimilarity measures have been proposed for financial time series clustering. These dissimilarity measures are used to cluster time series belonging to 3 data sets. One data set consists of EOD stock prices of 28 Indian companies. The second data set consists of EOD stock prices of 468 companies listed in S&P500 Index. The third data set consists of EOD stock prices of companies listed in Dow Jones Industrial Average Index. Overall, the data sets consist of 526 companies which is a fairly large number. Clustering is done and it is shown that our proposed dissimilarity measures outperform existing dissimilarity measures.

**Acknowledgments.** Kartikay Gupta was supported by Teaching Assistantship Grant by Ministry of Human Resource Development, India.

## A Appendix

### A.1 Correlation Based Dissimilarity Measure (COR)

A simple dissimilarity measure for time series clustering is based on Pearson's correlation factor between time series  $X_T$  and  $Y_T$  given by

$$COR(X_T, Y_T) = \frac{\sum_{t=1}^T (X_t - m_X)(Y_t - m_Y)}{\sqrt{\sum_{t=1}^T (X_t - m_X)^2} \sqrt{\sum_{t=1}^T (Y_t - m_Y)^2}}$$

where  $m_X$  and  $m_Y$  are the average values of the time series  $X_T$  and  $Y_T$  respectively. The dissimilarity measure is then given by

$$Dissimilarity_{COR}(X_T, Y_T) = \sqrt{2(1 - COR(X_T, Y_T))}$$

For more information regarding this distance measure, one may refer to [9].

## A.2 Temporal Correlation Based Dissimilarity Measure (CORT)

As introduced by Douzal [12], the similarity between two time series is evaluated using first order temporal correlation coefficient [13] given by,

$$CORT(X_T, Y_T) = \frac{\sum_{t=1}^{T-1} (X_{t+1} - X_t)(Y_{t+1} - Y_t)}{\sqrt{\sum_{t=1}^{T-1} (X_{t+1} - X_t)^2} \sqrt{\sum_{t=1}^{T-1} (Y_{t+1} - Y_t)^2}}$$

The dissimilarity proposed by Douzal [12] modulates the ‘dissimilarity value’ between  $X_T$  and  $Y_T$  using the coefficient  $CORT(X_T, Y_T)$ . Specifically, it is defined as follows.

$$Dissimilarity_{CORT}(X_T, Y_T) = \phi_k[CORT(X_T, Y_T)] \times Dissimilarity(X_T, Y_T)$$

where  $\phi_k$  is an adaptive function given by,

$$\phi_k(u) = \frac{2}{1 + e^{ku}}, \quad k \geq 0 \quad (6)$$

and  $Dissimilarity(X_T, Y_T)$  refers to dissimilarity value computed using any of the available dissimilarity measures like Euclidean, DTW etc. In this paper, we choose DTW as the preferred dissimilarity measure. This is because DTW effectively takes into account slight shape distortions while calculating its dissimilarity measure value. For more details regarding DTW, readers are referred to [9, 11].

## References

1. Dose, C.: Clustering of financial time series with application to index and enhanced index tracking portfolio. *Phys. A* **355**, 145–151 (2005)
2. Mantegna, S.: *An Introduction to Econophysics Correlations and Complexity in Finance*. Cambridge University Press, Cambridge (1999)
3. Basalto, N., et al.: Hausdorff clustering of financial time series. *Phys. A* **379**, 635–644 (2007)
4. Saeed, T.: Effective clustering of time-series data using FCM. *Int. J. Mach. Learn. Comput.* **4**(2), 170–176 (2014)
5. Guan, J.: Cluster financial time series for portfolio. In: *Proceedings of the 2007 ICWAPR*, Beijing, China, 2–4 November 2007
6. Marti, A., Nielson, D.: Clustering financial time series: how long is enough? In: *Proceedings of the Twenty-Fifth IJCAI* (2016)

7. Aghabozorgi, S., Shirkhorshidi, A.S., Wah, T.Y.: Time-series clustering - a decade review. *Inform. Syst.* **53**, 16–38 (2015)
8. John, G.: k-shape: efficient and accurate clustering of time series. *SIGMOD Rec.* **45**(1), 69–76 (2016)
9. Montero, V.: TSclust: an r package for time series clustering. *J. Stat. Softw.* **62**(1), 1–43 (2014)
10. Murtagh, L.: Ward’s hierarchical agglomerative clustering method: which algorithms implement Ward’s criterion? *J. Classif.* **31**, 274–295 (2014)
11. Berndt, D.J., Clifford, J.: Using dynamic time warping to find patterns in time series. In: *KDD, Workshop*, pp. 359–370 (1994)
12. Chouakria, A.D., Nagabhushan, P.N.: Adaptive dissimilarity index for measuring time series proximity. *Adv. Data Anal. Classif.* **1**(1), 5–21 (2007)
13. Chouakria-Douzal, A.: Compression technique preserving correlations of a multivariate temporal sequence. In: *Advances in Intelligent Data Analysis*, pp 566–577. Springer, Heidelberg (2003)
14. Gavrilov, M., et al.: Mining the stock market: which measure is best ? In: *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD*, pp. 487–496 (2000)
15. Giorgino, T.: Computing and visualizing dynamic time warping alignments in R: the dtw package. *J. Stat. Softw.* **31**(7), 1–24 (2009)
16. Hoffmann, M., et al.: Estimation of the lead-lag parameter from non-synchronous data. *ISI/BS Bernoulli* **19**(2), 426–461 (2013)