JOURNAL OF EVENT, TOURISM AND HOSPITALITY STUDIES

http://e-journal.uum.edu.my/index.php/jeth

How to cite this article:
Leong K., & Sung A. (2023). Co-movement clustering: A novel approach for predicting inflation in the food and beverage industry. *Journal of Event, Tourism and Hospitality Studies*, 3, 1-21. https://doi.org/10.32890/jeth2023.3.1

CO-MOVEMENT CLUSTERING: A NOVEL APPROACH FOR PREDICTING INFLATION IN THE FOOD AND BEVERAGE INDUSTRY

[1]Kelvin Leong & [2]Anna Sung
[1&2]Chester Business School, University of Chester, United Kingdom

[1]Corresponding author: k.leong@chester.ac.uk

Received: 23/2/2023 Revised: 4/8/2023 Accepted: 6/8/2023 Published: 31/10/2023

ABSTRACT

In the realm of food and beverage businesses, inflation poses a significant hurdle as it affects pricing, profitability, and consumer’s purchasing power, setting it apart from other industries. This study proposes a novel approach; co-movement clustering, to predict which items will be inflated together according to historical time-series data. Experiments were conducted to evaluate the proposed approach based on real-world data obtained from the UK Office for National Statistics. The predicted results of the proposed approach were compared against four classical methods (correlation, Euclidean distance, Cosine Similarity, and DTW). According to our experimental results, the accuracy of the proposed approach outperforms the above-mentioned classical methods. Moreover, the accuracy of the proposed approach is higher when an additional filter is applied. Our approach aids hospitality operators in accurately predicting food and beverage inflation, enabling the development of effective strategies to navigate the current challenging business environment in hospitality management. The lack of previous work has explored how time series clustering can be applied to support inflation prediction. This study opens a new research paradigm to the related field and this study can serve as a useful reference for future research in this emerging area. In addition, this study work contributes to the data analytics research stream in hospitality management literature.

Keywords: Inflation Rate Prediction, Clustering Technique, Time Series Analysis, Food and Beverage, Hospitality Industry.

INTRODUCTION

Inflation has become a global issue and is impacting many business sectors. Based on a survey conducted by JP Morgan (2022) involving more than 1,500 midsize business leaders, 71% of the responses indicated that inflation was their top challenge. This finding was echoed by another survey conducted by Babson College and David Binder Research (Goldman Sachs, 2022). The survey involved 1,533 small business owners in which 97% of small business owners today posited those inflationary pressures on their business had increased or stayed the same. Moreover, a similar message can also be found in a 2022 summer CEO Survey (Deloitte, 2022). The finding indicated that more than 80% of CEOs expected inflation to influence or disrupt their business strategy.

Food and Beverage (F&B) is an important area in the hospitality industry. Hotstats (2022) found that F&B revenue per available room in Europe increased by 155% from January to May 2022, reaching €42 in May 2022. In contrast, F&B revenue in the US lagged behind pre-pandemic levels but has increased by 117% since January 2022. However, recent rise in F&B price volatility and inflation led to a practical challenge for related hospitality businesses. Recent reports have indicated that restaurants, pubs, and bars in the UK were struggling to manage costs and attract footfall under the food cost inflation (McAllister, 2022), while restaurant owners in the US Bay Area also expressed that they had been negatively affected by the rise in food prices (Ramirez, 2022). In brief, garnered considerable attention and had a significant impact on the food and beverage industry. Inflation has a direct impact on food prices, the type, quantity, and quality of foods served in restaurants, and the amount of money establishment owners are willing to spend (Kubatko et al., 2023; Ali et al., 2022; D'Acunto et al., 2021; Diazgranados, 2022).

Although inflation has become a complex global issue that is beyond the control of individual businesses, effective F&B management can assist hospitality businesses in surviving the current harsh environment and gaining. This study focuses on a specific inflation management issue, which is to identify which F&B items are likely to inflate together. In practice, not all items inflate at the same time. These timing differences allow businesses to redesign their business plan agilely. For example, a restaurant chain could redesign its promotion plan by focusing on the food items which are not expected to inflate.

There are many ways to help businesses to predict what items will inflate but the point is, very often the information is incomplete. A business may be able to identify one or a few to-be-inflated items but on the other hand, does not realize some other items that will inflate together. In this regard, a business that can better predict what items will inflate together, will gain better information to make the best decision in cost saving, thus, a better competitive advantage is gained. Consider two hypothetical eateries, Fortunate Company and Unfortunate Company, both of which feature the same three meats (pork, beef, and chicken) on their menus. The price of pork is going up next month, according to both companies' suppliers. Fortunate Company uses this supplier data to their advantage by featuring beef throughout their menu. However, Unfortunate Company advertises chicken and has a high volume of chicken sales. Both pork and chicken prices will increase in the coming month. As a result, the inflated price of chicken had no effect on Fortunate Company, while Unfortunate Company lost money as a result of its promotion of chicken. In reality, every firm strives to attain the status of "Fortunate Company," which is synonymous with impeccable judgment. The difficulty of making good choices lies in foreseeing which goods will experience simultaneous inflation.

For the purpose of overcoming this challenge, a novel clustering approach is proposed namely, "co-movement clustering". The approach is to predict which items will inflate together according to historical time-series data. With the development of technology, more and more data are available in digital forms from various sources, such as digital accounting records or supply-chain records. These data can be easily organized into time series data. Time series data, also called time-stamped data, is a series of data items obtained through repeated measurements over regular time. The history of time series dates back to the 17th century, and it has been suggested that John Graunt was the inventor of modern time series (Klein, 1997). Time series analysis is useful because it can help users to foresee by learning how a given variable changes over time in the past.

In this study, the clustering approach is applied to learn from historical time series data and then to predict which items are likely to inflate together. In brief, clustering is the process of grouping objects into groups according to their similarity. The goal of clustering is to make the objects within a group similar to one another and different from the objects in other groups. There are several ways to conduct clustering and a more detailed discussion is provided in Section 2 on Related Works.

To date, lack of previous work has explored how time series clustering can be applied to support related inflation prediction. Accordingly, this study makes multiple contributions. First, it contributes to and extends the research stream related to inflation prediction by focusing on how to predict inflated-together items. Second, this study focuses on applying time series clustering to support inflation prediction which is a research gap in existing literature. Third, a novel clustering approach is proposed, namely co-movement clustering, whereby the prediction suggested by our approach based on real-world data is evaluated. The evaluation results indicate that the proposed approach outperformed classical clustering approaches. Therefore, this study opens a new research paradigm to the related field and serve as a useful reference for future research in this emerging area. Fourth, in this study, practical data analytics solutions are enriched to support businesses to overcome the inflation challenge. Fifth, this current work contributes to the data analytics research stream in hospitality management literature by suggesting a novel approach with experiment results as support.

This paper is structured as follows; Section 2 reviews the literature related to the proposed approach, thereby offering a rationale for the design of the approach. Section 3 explains the design of the proposed approach. Then, Section 4 of the study elucidates the application of the proposed co-movement clustering technique in facilitating inflation prediction within the food and beverage industry. Finally, Section 5 of this study presents the primary findings, expounds upon the theoretical and managerial contributions, and outlines the limitations of the research as well as potential avenues for future investigation.

LITERATURE REVIEW

This study proposes a time-series clustering approach to help businesses to predict which items will inflate together. Related work discussing on F&B cost prediction methods, time series, clustering, time series clustering, and similarity measures are explained as follows.

Food and Beverage Cost Prediction Methods

As a result of the rapid advancements in technology, numerous innovative methods and techniques have emerged in recent years to aid in the prediction of food and beverage costs. For example, Venkateswara et al. (2022) introduced a model for predicting the fluctuating prices of essential food materials, while Gao et al. (2022) presented a forecasting model specifically designed for agricultural price prediction. Watson et al. (2021) suggested the utilization of intelligent sensors to analyze food and drink usage in manufacturing processes. Other than that, Atalan (2023) proposed a novel algorithm to support the forecasting of milk prices, and Murugesan and Radha (2023) suggested a new model that utilizes hybrid ensemble learning techniques for predicting crop prices. Additionally, Sarangi et al. (2021) discussed a machine-learning approach for predicting the consumer food price index. However, there is a lack of related studies that combine time series analysis and clustering to support the prediction of food and beverage costs.

Time Series

Time series is adopted in this study because it effectively summarizes the price changes of an item over time. Therefore, such summary provides a meaningful reference for predicting the price change of the item in the future.

As per Oliveira and Antunes (2001), a time series is a temporal sequence that consists of continuous, real-valued elements (Aghabozorgi et al., 2015), and each time series can also be considered as a single object (Kumar & Nagabhushan, 2006). Furthermore, time series can be univariate, bivariate, or multivariate. In this study, time series refers to univariate time series which is the simplest form of temporal data and is a sequence of real numbers collected regularly, where each number represents a value (Wang et al., 2004).

Clustering

Clustering is a process of grouping items into groups according to their similarity. This study aims to predict which items will inflate together. In other words, this study aims to cluster food and beverage items into two groups: i) items inflated together; and items ii) items not inflated together. Therefore, clustering is applied in this study.

Major clustering methods can be classified as partition-based, hierarchical, density-based, grid-based, and statistical model-based methods. Partition-based clustering is useful when the number of clusters is known, and this method produces sphere-like clusters. Typical partition-based clustering examples include K-means (MacQueen, 1967; Forgy, 1965), CLARA (Kaufman & Rousseeuw, 2009), and CLARANS (Ng & Han, 2014). Among these approaches, K-means is considered the most common approach. Hierarchical clustering is further subdivided into agglomerative or divisive approaches – both approaches build trees of clusters. Some examples of hierarchical clustering methods include CURE (Guha et al., 1998), BIRCH (Zhang et al., 1996), and Chameleon (Karypis et al., 1999). Density-based clustering including DBSCAN (Ester et al., 1996) and OPTICS (Ankerst et al., 1999) was developed to handle arbitrary-shaped clusters. Grid-based clustering such as STING (Wang et al., 1997), WaveCluster (Sheikholeslami et al., 1998) and CLIQUE (Agrawal et al., 1998) produce clusters based on grids; this type of method quantizes the object space into a finite number of cells that form a grid structure. The probabilistic model-based methods such as the EM algorithm (Dempster et al., 1977) and AutoClass (Cheeseman & Stutz, 1996) assume that the data come from a mixture model of several populations (McLachlan & Basford, 1988).

In application, this study clusters items into two clusters: one cluster is items that inflate together, and another cluster is items do not inflate together. Given the predetermined number of clusters in the proposed work, partition-based clustering is adopted.

Time Series Clustering

Time series clustering is a special type of clustering. In general, time series clustering can be used for various purposes, such as detection of correlation (Gopikrishnan et al., 2000; He et al., 2012), finding interesting patterns (Aljawarneh et al., 2016), and prediction and recommendation (Graves & Pedrycz, 2010; Ito et al., 2009; Pavlidis et al., 2006; Sfetsos & Siriopoulos, 2004). In brief, time series clustering can be achieved by (i) whole time series clustering, (ii) subsequence clustering, and (iii) time point clustering. For the first approach, whole time series clustering is conducted based on a set of individual time-series data with respect to their similarity: the basic unit is time-series. The second approach, subsequence clustering involves the clustering of a set of subsequences (i.e., segments) from a single long time series. The third approach is time point clustering which refers to the clustering of time points according to their temporal proximity and the similarity of their corresponding values.

In this study, the first approach is adopted – whole time series clustering, in which the whole time series of each item was used as the basic unit for clustering. In brief, the whole time series contains useful information about how the price of an item changed over time. Therefore, this approach provides a meaningful reference for supporting prediction on price change. More specifically, this study clusters items according to the similarity of historical price change of each item into the two predefined clusters. A similarity measure as an evaluation measure is discussed in the next section.

Similarity Measure

Very often, the similarity between objects is measured by distance in clustering. Therefore, distance measure plays an important role in the similarity measure.

There are different ways to measure the distance between two-time series, such as correlation (Berthold and Höppner, 2016; Focardi & Fabozzi, 2001), Euclidean distance (Chen, 2005; Kalpakis et al., 2001), Dynamic time warping (DTW) (Anh & Thanh, 2015; Guoqing et al., 2001), and cosine similarity (Wu et al., 2018; González-Vidal et al., 2018). These measures are discussed as follows:

Although correlation does not measure similarity, correlation can determine the relationship between variables from two-time series and the strength of that relationship. Therefore, correlation has been used in time series clustering in previous studies such as in Berthold and Höppner, (2016); Siyou Fotso et al., (2020); Egri et al., (2017). On the other hand, Euclidean distance refers to finding out the distance between two points in Euclidean space. It essentially represents the shortest distance between two points. Euclidean Distance is one of the most used distance metrics. Among these measures, a previous study by Lkhagva et al. (2006), suggested that the Euclidean distance is surprisingly competitive in terms of time-series classification accuracy. However, if two data vectors have no attribute values in common, by using Euclidean distance, they may have a smaller distance than the other pair of data vectors containing the same attribute values (Jain et al., 1999). Consequently, various similarity measures have been proposed to overcome this limitation, one of these is called Dynamic Time Warping (DTW). DTW is a time series alignment algorithm which was originally developed for speech recognition (Sakoe & Chiba, 1978). In time series application, the main idea of DTW is to compute the distance from matching similar elements between time series. It uses the dynamic programming technique to find the optimal temporal matching between elements of two-time series. The main feature of the DTW distance measure is that it allows recognition of similar shapes, even if they present signal transformations, such as shifting or scaling. Alternatively, Cosine similarity measures the similarity by comparing the angles between two vectors and determines whether two vectors are pointing in roughly the same direction. The smaller the angle, the higher the similarity. Cosine similarity's disadvantage is even if two similar objects are far apart by Euclidean distance, they could still have a smaller angle between them, i.e., less distant.

Although the above measures have been widely used in the evaluation of results of time series clustering, they consider how time series were co-moved together over time. In this regard, we consider co-movement to be a key factor in clustering to identify which items will inflate together. Therefore, a novel measure is proposed in this study. Moreover, the above-mentioned four classical similarity measures were used as threshold and their prediction results were compared against our proposed approach.

THE DESIGN OF CO-MOVEMENT CLUSTERING

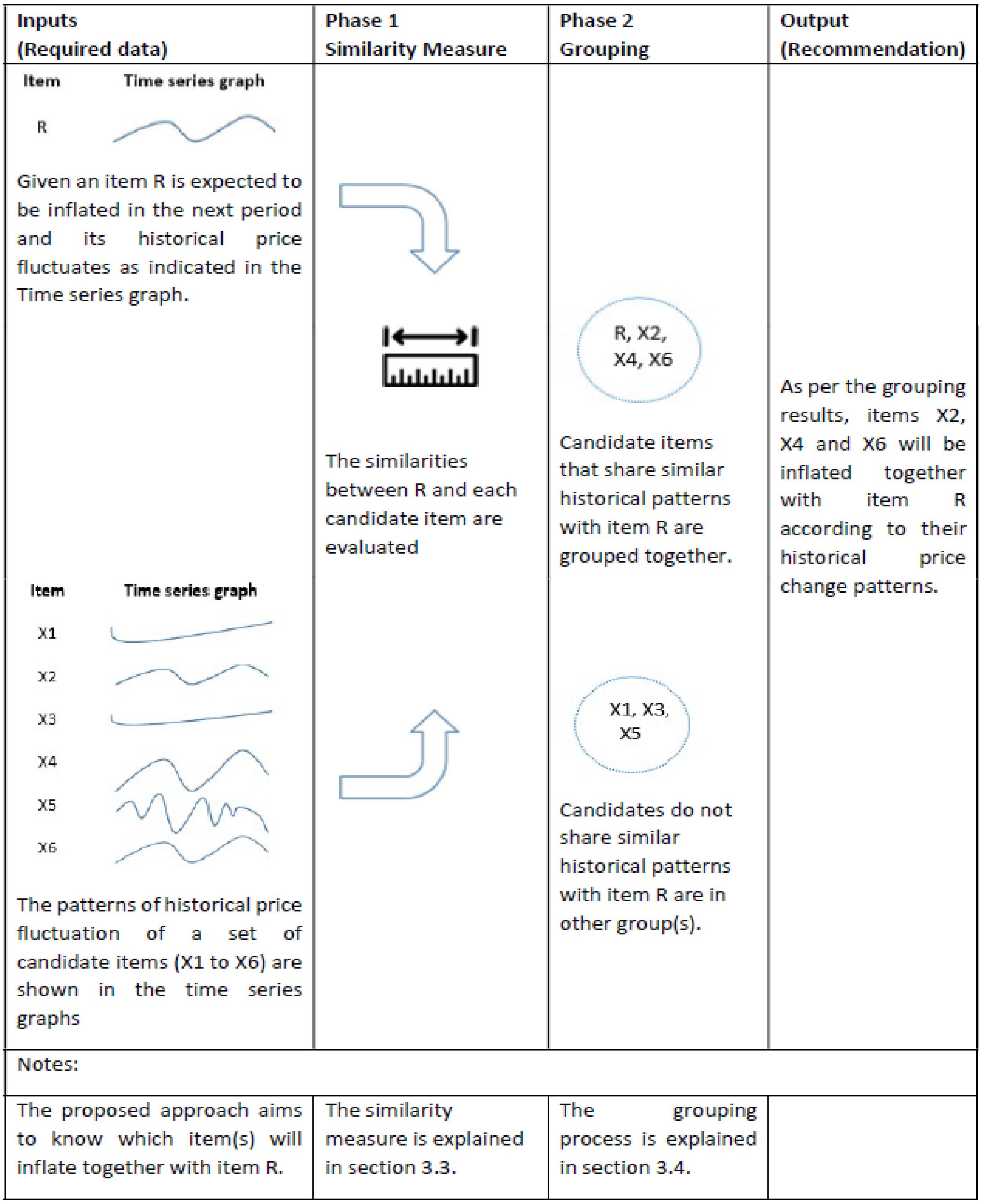
In this section, the design of the proposed co-movement clustering and its underlying concept is explained.

The Concept Behind Co-Movement Clustering

Co-movement clustering is designed based on the concept below:
It is assumed that history often repeats itself. Based on this assumption, if the patterns of price change of two products are similar in the past, their patterns of price change will continue or at least look similar in the future. Therefore, a novel approach is proposed in this study to predict which products will inflate together according to the patterns of products' historical price changes. Figure 1 illustrates the design of the proposed approach.

Figure 1

Design of the Proposed Approach



The design of the proposed approach is explained in more detail in Section 3.2 and a set of simplified example data is used to support the explanation.

The Use of Example Data in the Proposed Approach

The example data consists of a set of time series as shown in Table 1 and Figure 2.

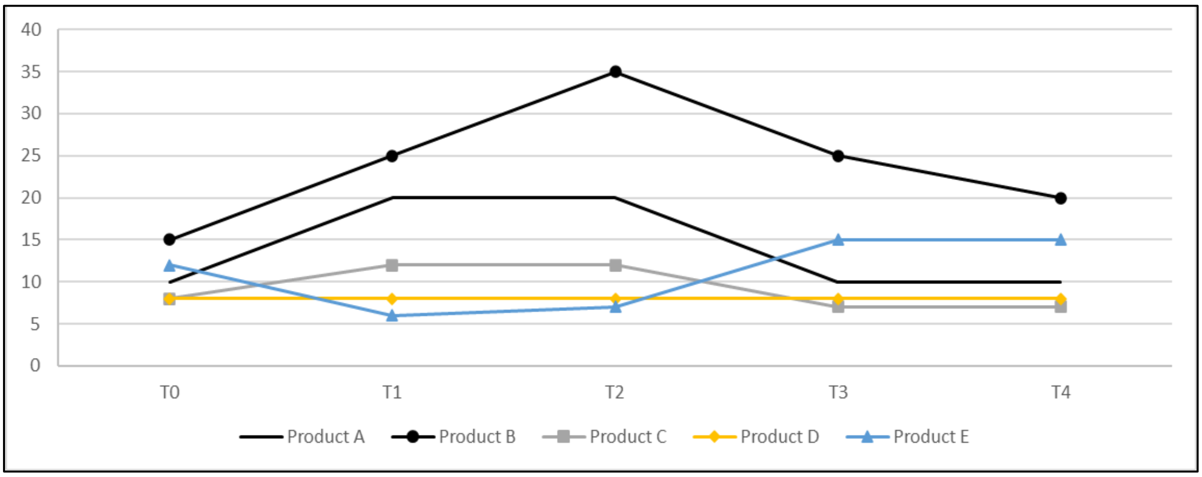
Table 1

Example of Time Series Dataset for Products A to E During a Period from T0 to T4

Observation timestamp:	T0	T1	T2	T3	T4
Time series of Product A	10	20	20	10	10
Time series of Product B	15	25	35	25	20
Time series of Product C	8	12	12	7	7
Time series of Product D	8	8	8	8	8
Time series of Product E	12	6	7	15	15

Figure 2

Time Series of Product A to Product E



Assuming it is expected that product A is going to inflate in the next period, T5, our interest is to identify what other products (i.e., B, C, D and E) will also inflate in T5 together with product A.

The proposed approach uses clustering to identify a group of to-inflate products according to the similarity of products' historical time series patterns. The identification process involves two phases: Phase (1) similarity measure and Phase (2) grouping. These two phases are explained in the following sub-sections.

Phase 1: Similarity Measure

The term 'reference item' in this study is used to indicate the item which is expected to inflate while other items are termed as "candidates". Therefore, in this case, product A is the reference item while products B, C, D and E are the candidates.

In phase 1, the proposed approach aims to measure the similarity between the reference item and the candidates in terms of their time series patterns.

To achieve this aim, a distance measure is proposed to measure the similarity between the reference item (product A) and each of the candidates (product B, C, D, and E). The distance measure is obtained by counting how many times a pair of time series moves in the same direction and then dividing the count by total observation.

For example, when we compare the time series of two products: product A and product B from T0 to T4 as per Table I, according to the figures in the table, products A and B are considered as moving in the same direction in T1 (both inflated from T0) and T3 (both deflated from T2). However, the movement direction in T2 and T4 is considered not in the same direction. Overall, the similarity measure of these two-time series of products A and B, from T0 to T4 equals 0.5 (2 times moving in the same direction divided by 4 observations: T1, T2, T3, and T4.) By using the same approach, the similarity measures between products C, D, and E to product A are 1, 0.5, and 0.25, respectively.

Phase 2: Grouping

Phase 2 of the proposed approach is grouping. Grouping aims to cluster candidates (products B, C, D, and E) into different groups according to their similarities with the reference item (product A) through a partition-based method.

Given the purpose of the application in this study is to identify items that are inflated together with the reference item, therefore we suggest the number of clusters to be equal to two: one cluster refers to the inflated together group and the second cluster refers to the not inflated together group. Accordingly, we adopt a mid-point approach to classify time series into two groups. The midpoint of the distance measure is the middle of all calculated similarity measures. Taking our scenario as an example, given the calculated maximum similarity measure between products B, C, D, and E to product A is 1 (i.e., product C) and the calculated minimum similarity measure is 0.25 (i.e. product E), therefore the mid-point is 0.625 (i.e. (1+0.25)/2).

Once the mid-point is obtained, candidates are then classified into two groups, for the items with similarity measures greater than the mid-point, the items are classified as an inflated together group, otherwise, the items are classified into non-inflated together group. It is worth mentioning that the above two clusters grouping approach is proposed according to the context of our case, which is to identify which items inflate together with the reference item. Other grouping approaches could also be considered in other applications according to corresponding contexts.

THE APPLICATION OF CO-MOVEMENT CLUSTERING

This section demonstrates how the proposed co-movement clustering can be used to support inflation prediction by F&B businesses. We also compared the accuracy of the proposed approach against the selected threshold similarity measures/methods.

Data

The first step to using the proposed approach is to identify a list of F&B items. In this demonstration, 14 Consumer Price Inflation (CPI) items obtained from the ONS (Office for National Statistics, UK.) were selected for the purpose. In practice, F&B businesses can select a list depending on the availability of data. A key requirement for the data is that the data should be obtained through repeated measurements over time, that is, in a time series format.

For the 14 CPI items, they are presented on a monthly basis in a time series format. We collected relevant data from January 1988 (i.e., the first month with relevant available data) to June 2022, in other words, a dataset with 5,782 data (i.e., 14 items x 413 months [from Jan 1988 to June 2022]) was constructed for the purpose.

These 14 CPI items are the full list of products under the same food and beverage category level in the ONS list, therefore we consider the selection of this dataset from ONS to be relevant, fair, and representative. In addition, given this dataset is available online, the results of this experiment are not only replicable but further studies can also be conducted based on the same or extended dataset to provide comparable results.

Training Data

The function of training data is for the proposed approach to learn how the identified items behaved in the past so that the approach can predict how the items would likely change in the future.

Given that we demonstrated the usefulness of the proposed method for inflation forecasting with six attempts between January 2022 and June 2022, the entire dataset was divided into six training data sets as follows:

Table 2

List of Periods of Training Data and Prediction Data

Training data (period)	Inflation prediction attempt (period)
January 1988 to December 2021	January 2022
January 1988 to January 2022	February 2022
January 1988 to February 2022	March 2022
January 1988 to March 2022	April 2022
January 1988 to April 2022	May 2022
January 1988 to May 2022	June 2022

Reference Item

As explained in the previous section, a reference item refers to the “to-inflate” item identified by a business. In practice, business decision makers (e.g., a café manager) may have some knowledge about inflation trends and they may be able to identify one or some of the “to-inflate” items from different sources, such as trade magazines, news, and conferences.

In this demonstration, we used the item with the highest monthly increase rate in the current month as the reference item, and then the data from previous months were used as the training set in the proposed approach to predict what are the inflated-together items.

Evaluation Design

By using the training data and reference item, the proposed co-movement approach was utilized to predict which other items will inflate-together with the reference item. The accuracy was evaluated according to the prediction result. As explained above, the last six months of the period were selected (i.e., from January 2022 to June 2022) for the evaluation purpose. For more detail, Section 4.2 serves as an example to illustrate the evaluation process.

Evaluation Measure

An accuracy index was used to evaluate the accuracy of the approach. The index is calculated as below:

$$\text{Accuracy index} = \frac{Sum\ of\ corrected\ suggested\ direction}{Total\ number\ of\ suggestions\ for\ the\ direction} \tag{1}$$

There were two possible directions as outcomes of each item monthly: (a) inflated and (b) non-inflated. For example, assuming the real directions are product A inflated and product B non-inflated, the accuracy index would be 1 if the approach suggested the same, while the accuracy index would be 0.5 only if the approach suggested both as inflated. The accuracy index ranges from 0 to 1, with higher values indicating greater precision. Other than the proposed co-movement approach, other classical similarity measures were also applied based on the same approach for comparison purposes. These classical similarity measures include Correlation, Euclidean distance, Cosine Similarity, and DTW.

CPI Data Description

As explained above, the proposed approach was utilized to predict which selected CPI items inflated-together during the selected six-month period from January 2022 to June 2022.

The CPI indexes by month of selected items from Jan 2022 to June are summarized in Table 3.

Table 3

CPI Indexes by Month of Selected Items from Jan 2022 to June 2022

CPI items	2022 JAN	2022 FEB	2022 MAR	2022 APR	2022 MAY	2022 JUN
01-BREAD & CEREALS	108.30	110.10	110.00	112.3	114.1	116.00
02-MEAT	103.10	103.80	104.20	106.2	107.5	110.30
03-FISH	114.30	115.70	116.20	118.5	118.1	121.20
04-MILK, CHEESE & EGGS	104.80	107.60	108.40	110.3	112.5	116.10
05-OILS & FATS	126.70	123.20	132.10	130.8	135.6	136.50
06-FRUIT	116.70	117.20	115.70	115.9	115.9	116.50
07-VEGETABLES INCLUDING POTATOES AND OTHER TUBERS	105.40	106.60	105.70	107.1	109.4	110.40
08-SUGAR JAM HONEY SYRUPS CHOCOLATE & CONFECTIONERY	106.80	105.40	105.10	107.2	109.2	107.00
09-FOOD PRODUCTS	108.50	110.50	109.90	113.4	114.1	118.20
10-COFFEE, COCOA	106.50	107.40	107.20	108.6	111.8	111.60
11-MINERAL WATERS, SOFT DRINKS AND JUICES	112.90	115.00	116.70	116.4	118.9	117.10
12-SPIRITS	101.40	100.40	101.70	101.9	102.9	102.00
13-WINE	105.40	104.10	105.10	104.3	104.8	105.50
14-BEER	107.50	109.20	109.60	108.9	109.5	109.90

As per Table 3, the monthly CPI indexes changed at different rates during the period. Table 4 summarizes the monthly change of the CPI indexes during the selected six-month period. According to the table, among the 14 items, 4 items had achieved the highest monthly changes during the period, including i) milk, cheese & eggs (in Feb 2022), ii) oils & fats (in Mar 2022 and May 2022), iii) food products (in Apr 2022 and Jun 2022) and iv) spirits (in Jan 2022). Overall, inflations were found 61 times for these 14 items during the period and meat was the only item that continuously inflated every month.

Table 4

Monthly Change (in percentage) of CPIs among Selected Items

CPI items	2022 JAN Change	2022 FEB Change	2022 MAR Change	2022 APR Change	2022 MAY Change	2022 JUN Change
01-BREAD & CEREALS	-0.64%	1.66%	-0.09%	2.09%	1.60%	1.67%
02-MEAT	0.29%	0.68%	0.39%	1.92%	1.22%	2.60%
03-FISH	2.24%	1.22%	0.43%	1.98%	-0.34%	2.62%
04-MILK, CHEESE & EGGS	-0.47%	2.67%	0.74%	1.75%	1.99%	3.20%
05-OILS & FATS	2.26%	-2.76%	7.22%	-0.98%	3.67%	0.66%
06-FRUIT	0.69%	0.43%	-1.28%	0.17%	0.00%	0.52%
07-VEGETABLES INCLUDING POTATOES AND OTHER TUBERS	0.48%	1.14%	-0.84%	1.32%	2.15%	0.91%
08-SUGAR JAM HONEY SYRUPS CHOCOLATE & CONFECTIONERY	3.19%	-1.31%	-0.28%	2.00%	1.87%	-2.01%
09-FOOD PRODUCTS	-0.09%	1.84%	-0.54%	3.18%	0.62%	3.59%
10-COFFEE, COCOA	2.11%	0.85%	-0.19%	1.31%	2.95%	-0.18%
11-MINERAL WATERS, SOFT DRINKS AND JUICES	2.26%	1.86%	1.48%	-0.26%	2.15%	-1.51%
12-SPIRITS	3.47%	-0.99%	1.29%	0.20%	0.98%	-0.87%
13-WINE	1.93%	-1.23%	0.96%	-0.76%	0.48%	0.67%
14-BEER	1.22%	1.58%	0.37%	-0.64%	0.55%	0.37%

Performance of Inflation Prediction

The proposed co-movement approach was used to predict which CPI items inflated-together for every month during the selected period. More specifically, the proposed approach was used to classify items into two clusters: (i) inflated-together items and (ii) not-inflated-together items. First, the item with the highest monthly increase rate in every current month was selected as the reference item and then the data from the corresponding previous months were treated as a training set to calculate the similarity between the reference item and each of the other items (candidates). For example, in the current experiment, knowing that oil and fat had the highest monthly increase rate among all other CPIs in May 2022, it is identified as the reference item and then the data from the previous months (i.e. from January 1988 to April 2022) were fed as the training data in the proposed approach to predict the inflated-together items in May 2022.

Similarity measures were calculated for each candidate item during the period. The method of calculation is explained in Section 3.3. After the similarity measures were obtained for each candidate item, a mid-point approach was adopted to classify items into two clusters. The evaluation measure as explained in the beginning of Section 4 was used to evaluate the accuracy. Table 5 summarizes the performance of the proposed approach. Overall, the measures of accuracy ranged from 38.46% to 69.23% over the selected period.

Table 5

Performance Measures of the Proposed Approach in Terms of Accuracy.

Correctly identified items by the proposed approach	Direction of change	Accuracy (correct prediction)
Jan 2022 (Reference item = 12-SPIRITS)		
08-SUGAR JAM HONEY SYRUPS CHOCOLATE &	Inflated	**53.85%**
CONFECTIONERY	Inflated	
11-MINERAL WATERS, SOFT DRINKS AND JUICES	Inflated	
13-WINE	Inflated	
14-BEER		
01-BREAD & CEREALS	Deflated	
04-MILK, CHEESE & EGGS	Deflated	
09-FOOD PRODUCTS	Deflated	
Feb 2022 (Reference item = 04-MILK, CHEESE & EGGS)		
01-BREAD & CEREALS	Inflated	**69.23%**
02-MEAT	Inflated	
06-FRUIT	Inflated	
07-VEGETABLES INCLUDING POTATOES AND	Inflated	
OTHER TUBERS	Inflated	
11-MINERAL WATERS, SOFT DRINKS AND JUICES		
05-OILS & FATS	Deflated	
08-SUGAR JAM HONEY SYRUPS CHOCOLATE &	Deflated	
CONFECTIONERY	Deflated	
12-SPIRITS	Deflated	
13-WINE		
Mar 2022 (Reference item = 05-OILS & FATS)		
02-MEAT	Inflated	**38.46%**
13-WINE	Inflated	
14-BEER	Inflated	
06-FRUIT	Deflated	
10-COFFEE, COCOA	Deflated	
Apr 2022 (Reference item = 09-FOOD PRODUCTS)		
01-BREAD & CEREALS	Inflated	**38.46%**
02-MEAT	Inflated	
07-VEGETABLES INCLUDING POTATOES AND	Inflated	
OTHER TUBERS		
05-OILS & FATS	Deflated	
13-WINE	Deflated	

May 2022 (Reference item = 05-OILS & FATS)		
01-BREAD & CEREALS	Inflated	**69.23%**
02-MEAT	Inflated	
07-VEGETABLES INCLUDING POTATOES AND	Inflated	
OTHER TUBERS	Inflated	
08-SUGAR JAM HONEY SYRUPS CHOCOLATE &	Inflated	
CONFECTIONERY	Inflated	
09-FOOD PRODUCTS	Inflated	
13-WINE		
14-BEER		
03-FISH	Deflated	
06-FRUIT	Deflated	
Jun 2022 (Reference item = 09-FOOD PRODUCTS)		
01-BREAD & CEREALS	Inflated	**53.85%**
02-MEAT	Inflated	
07-VEGETABLES INCLUDING POTATOES AND	Inflated	
OTHER TUBERS		
08-SUGAR JAM HONEY SYRUPS CHOCOLATE &	Deflated	
CONFECTIONERY	Deflated	
10-COFFEE, COCOA	Deflated	
11-MINERAL WATERS, SOFT DRINKS AND JUICES	Deflated	
12-SPIRITS		

In addition, the performance was compared to that of four traditional techniques (correlation, Euclidean distance, Cosine Similarity, and DTW). As shown in Table 6, the proposed method outperformed other methods for four out of six trials, and its average accuracy index (0.5385) performs better than all other traditional methods. Therefore, we proposed that the proposed method provides a more precise prediction of inflation among items.

Table 6

Comparison of Prediction Accuracy of the Proposed Approach and Classical Methods

	Correlation	Euclidean distance	Cosine Similarity	DTW	The proposed approach
Jan-22	0.4615	0.3846	0.3846	0.3846	0.5385*
Feb-22	0.6154	0.6154	0.6154	0.5385	0.6923*
Mar-22	0.3077	0.3846	0.3846	0.4615*	0.3846
Apr-22	0.6154*	0.6154*	0.5385	0.5385	0.3846
May-22	0.6154	0.5385	0.5385	0.3077	0.6923*
Jun-22	0.3077	0.4615	0.5385*	0.3846	0.5385*
In Average	0.4872	0.5000	0.5000	0.4359	0.5385*

Further Improvement

Although the proposed approach has been demonstrated as a more accurate approach than other classical methods in the above cases, the results can be further improved by using the calculated similarity measures as a filtering criterion.

As shown in Table 7, when a filtering criterion is set to include only items with a similarity measure greater than 0.5, the output (recommendation) contains only those items. The predicted results match the actual results in all of the months during the chosen period.

Table 7

Performance Measures of the Proposed Approach in Terms of Accuracy (with Similarity Measure > 0.5)

Clustering results (output = recommendation)	Predicted direction	Actual direction	Accuracy (correct prediction)
In Jan-22			100%
13-WINE	Inflated	Inflated	
14-BEER	Inflated	Inflated	
08-SUGAR JAM HONEY SYRUPS CHOCOLATE & CONFECTIONERY	inflated	inflated	
In Feb-22	Inflated	Inflated	100%
01-BREAD & CEREALS	Inflated	Inflated	
07-VEGETABLES INCLUDING POTATOES AND	Inflated	Inflated	
OTHER TUBERS	Inflated	Inflated	
06-FRUIT			
02-MEAT			
In Mar-22	Inflated	Inflated	100%
02-MEAT			
In Apr-22	Inflated	Inflated	100%
02-MEAT	Inflated	Inflated	
07-VEGETABLES INCLUDING POTATOES AND	inflated	inflated	
OTHER TUBERS			
01-BREAD & CEREALS			
In May-22	Inflated	Inflated	100%
02-MEAT			
In Jun-22	Inflated	Inflated	100%
02-MEAT	Inflated	Inflated	
07-VEGETABLES INCLUDING POTATOES AND	Inflated	Inflated	
OTHER TUBERS			
01-BREAD & CEREALS			

DISCUSSION AND CONCLUSION

Cost prediction plays a critical role in the F&B industry as it directly influences pricing strategies and supply chain management. Accurate predictions enable businesses to set competitive menu prices while maintaining profitability, leading to long-term sustainability. Additionally, cost prediction aids in effective supply chain management by informing procurement decisions, optimizing inventory levels, and mitigating risks associated with price fluctuations. By leveraging cost prediction, businesses can enhance their financial performance, maintain market competitiveness, and foster a resilient and efficient food and beverage industry. Inflation has become a key challenge for many food and beverage businesses. Therefore, an approach that can more accurately identify which items inflated together can better support those businesses to make better decisions.

This study has suggested a novel approach, namely co-movement clustering, which uses historical time series data to predict which items are likely to-inflate together. Our experiment results demonstrated that its accuracy outperforms other classical methods (correlation, Euclidean distance, Cosine Similarity, and DTW). Moreover, the accuracy is higher when an additional filter is applied. It is noteworthy to mention that this study is centered on the analysis of data specifically from the UK. However, it is important to acknowledge that different countries may utilize distinct primary ingredients within their respective food and beverage industries. Nevertheless, the analytical outcomes possess the potential for generalization across diverse countries.

Theoretical Implications

This study offers several theoretical contributions to hospitality research. The first contribution is on similarity measures. In practice, similarity can be determined from different points of view, however, although many similarity measures had been proposed in previous studies, only very few of them had been applied with a focus on evaluating the co-moving patterns among time series. This study introduces a novel measure derived from the co-moving pattern of time series perspective to determine the similarity between food and beverage items. The results of our demonstration based on real-world CPI data proved that the proposed approach outperformed other classical similarity measures, such as correlation, Euclidean distance, Cosine Similarity, and DTW in terms of inflation prediction accuracy. From our point of view, the reason why the proposed co-moving measure can generate more accurate results was because the proposed measure takes the moving direction of the time series over regular time points into consideration. This novel but practical and make-sense approach allows us to better predict which food and beverage items will inflate together. In fact, this study opens a new research arena of similarity measures on time series, applicable in the hospitality industry.

The second contribution is in terms of its methodological approach. This study proposes a novel approach to identify inflated-together food and beverage items by using time series clustering. Time series summarizes how the price of a food and beverage item changes over time by encapsulating the underlying causes and factors of relevant information into systemic patterns. Given history always repeats itself, therefore, time series can serve as an effective proxy that helps to predict potential patterns of inflation in the future. In addition, partitional clustering was used as an unsupervised learning approach to classify time series into two groups (i.e., inflated together and not inflated together) based on their similarity of historical pattern. By using an unsupervised learning approach, hidden patterns can be determined which are not possible using traditional methods such as visualization. In brief, the proposed approach combines unsupervised learning with effective data proxy (i.e., time series) and the findings proved the accuracy of this novel approach. More importantly, the approach can serve as a useful reference for future research in terms of how to use time series clustering in supporting inflation prediction.

The third contribution is on exploring the use of unsupervised data analytics to support food and beverage inflation prediction. Although inflation has been widely studied in the field of social science research, most relevant studies were hypothesis-driven, that is, those studies were driven by hypotheses derived from theories and then through data collection and analysis to validate the hypotheses. In this study, a novel way to support inflation prediction has been introduced, different from a hypothesis-driven approach. In other words, this study attempts to use data to determine how a product's price will change according to historical patterns. The findings of our demonstration showed very accurate result (100% accurate when a 50% filter was applied). In brief, this current work establishes a new research arena for this topic.

Practical Implications

Given that food and beverage costs contribute significantly to the hospitality industry, it is crucial that food and beverage operators are able to make more accurate predictions regarding which food and beverage items will inflate together. This allows for the development of a more robust business strategy. If a café manager anticipates that the price of fish and beef will rise in tandem over the next month, she may choose to eliminate or reduce the number of fish and beef dishes available on the set menu for the following month. A cafe manager, for example, might be able to anticipate some of the "to-inflate" food and beverage items by consulting various sources like trade magazines, news articles, conferences, etc. The problem, however, is figuring out how to recognize the other unnamed food and drink items that will expand at the same time. In fact, the better a business plan can be made with the more items ‘to-inflate”. The paper presents a new approach to support businesses to tackle the above-mentioned challenge. We demonstrated how to use the proposed approach to predict inflated-together items based on real-world data from ONS (i.e., CPI items). In practice, the current approach can be easily adopted by food and beverage businesses. In order to use the proposed approach, food and beverage businesses can easily find historical price records to replace the CPI indexes. For example, a café manager would be able to obtain average purchasing monthly costs from their accounting records to develop a time series data.

In fact, this current approach can also encourage financial practitioners in the sector to gain additional insight from available historical records. In particular, with the increasing power of IT technology, data can be stored and kept for a long time, therefore, more food and beverage data would be ready in the form of time series data. In this regard, financial practitioners can use this proposed approach with available data to identify which food and beverage items will inflate together and then combine it with financial forecast techniques to forecast the potential impacts of different possible financial scenarios. For example, a management accountant could apply Cost-Volume-Profit (CVP) analysis based on the suggested results from this proposed approach to optimizing the product mix.

Limitations and Future Research

Most of the limitations of this study are data related. In this paper, the use of the proposed approach has been demonstrated based on CPI data from ONS, and the data is organized at the aggregated level. For example, one of the CPI categories is “fruit” which aggregates the prices of all types of fruits together. However, the prices of different types of fruit would change differently depending on many factors, such as origins, logistics, and seasons. Therefore, aggregated data may lead to some levels of information loss and the accuracy of this proposed approach based on disaggregated data should be further investigated. Moreover, given the purpose of the proposed approach is to identify inflate-together food and beverage items, therefore, the proposed approach is designed to cluster food and beverage items according to their historical patterns of price change direction, but the proposed approach does not consider the magnitude of price change. In other words, the proposed approach handles minor price changes and significant price changes in the same manner if the direction of price change is the same. As a result, the impact of magnitude change cannot be reflected. Furthermore, the performance of the proposed approach in terms of processing speed was not evaluated. Currently, the dataset used in the demonstration contains 5,782 data (i.e., 14 items x 413 months [from Jan 1988 to

June 2022]), and the size of the dataset is relatively small in the big data age. The processing speed performance of the proposed approach with a huge dataset can be evaluated in future studies.

Building on this study, several future research directions have been identified. Firstly, an analysis mechanism should be developed to evaluate the suggested inflate-together items generated by the proposed approach. The analysis mechanism should compare the suggestion against actual change to monitor the accuracy of the proposed approach. The comparison results could inform the performance of the proposed approach over time and on each item. As a result, it helps decision-makers to make better decisions. Secondly, as mentioned previously, the proposed approach does not take the magnitude of price change direction into consideration, future studies could incorporate magnitude as a consideration in the prediction. In addition, the present study places emphasis on the implementation of the suggested approach for forecasting inflation rates specifically pertaining to food and beverage products. Further investigation is warranted to broaden the scope of application to encompass diverse categories of items and to assess the viability of the proposed approach. The evaluation of the processing speed performance of the proposed approach can be conducted in a systematic manner by utilizing a large dataset.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENT

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

Aghabozorgi, S., Seyed Shirkhorshidi, A., & Ying Wah, T. (2015). Time-series clustering – A decade review. *Information Systems, 53,* 16-38. https://doi.org/https://doi.org/10.1016/j.is.2015.04.007

Agrawal, R., Gehrke, J., Gunopulos, D., & Raghavan, P. (1998). *Automatic subspace clustering of high dimensional data for data mining applications* Proceedings of the 1998 ACM SIGMOD International Conference on Management of data, Seattle, Washington, USA. https://doi.org/10.1145/276304.276314

Ali, M. H., Iranmanesh, M., Tan, K. H., Zailani, S., & Omar, N. A. (2022). Impact of supply chain integration on halal food supply chain integrity and food quality performance. *Journal of Islamic Marketing*, *13*(7), 1515-1534.

Aljawarneh, S., Radhakrishna, V., Kumar, P. V., & Janaki, V. (2016, 22-24 Sept. 2016). *A similarity measure for temporal pattern discovery in time series data generated by IoT*. 2016 International Conference on Engineering & MIS (ICEMIS).

Anh, D. T., & Thanh, L. H. (2015). An efficient implementation of k-means clustering for time series data with DTW distance. *International Journal of Business Intelligence and Data Mining*, *10*(3), 213-232. https://doi.org/10.1504/IJBIDM.2015.071311

Ankerst, M., Breunig, M. M., Kriegel, H.-P., & Sander, J. (1999). *OPTICS: Ordering points to identify the clustering structure* Proceedings of the 1999 ACM SIGMOD international conference on

Management of data, Philadelphia, Pennsylvania, USA. https://doi.org/10.1145/304182.304187

Atalan, A. (2023). Forecasting drinking milk price based on economic, social, and environmental factors using machine learning algorithms. *Agribusiness*, *39*(1), 214-241.

Berthold, M. R., & Höppner, F. (2016). On clustering time series using euclidean distance and pearson correlation. *arXiv:1601.02213*. https://doi.org/10.48550/ARXIV.1601.02213

Cheeseman, P., & Stutz, J. (1996). Bayesian classification (AutoClass): theory and results," in advances in knowledge discovery and data mining. In U. M. Fayyad, G. Piatetsky-Shapiro, S. P., & R. Uthurusamy (Eds.), *Advances in Knowledge Discovery and Data Mining* (pp. 153-180). MIT Press.

Chen, J. R. (2005, 2005). Making subsequence time series clustering meaningful. Fifth IEEE International Conference on Data Mining (ICDM'05).

D'Acunto, F., Malmendier, U., Ospina, J., & Weber, M. (2021). Exposure to grocery prices and inflation expectations. *Journal of Political Economy*, *129*(5), 1615-1639.

Deloitte. (2022). *Summer 2022 Fortune/Deloitte CEO Survey*. https://www2.deloitte.com/us/en/pages/chief-executive-officer/articles/ceo-survey.html

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, *39*(1), 1-22. https://doi.org/https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

Diazgranados, D. (2022). *3 Ways Inflation is Impacting the Food and Beverage Industry*. Entrepreneur. https://www.entrepreneur.com/money-finance/3-ways-inflation-is-impacting-the-food-and-beverage-industry/435900

Egri, A., Horváth, I., Kovács, F., Molontay, R., & Varga, K. (2017, 20-23 Oct. 2017). Cross-correlation based clustering and dimension reduction of multivariate time series. 2017 IEEE 21st International Conference on Intelligent Engineering Systems (INES).

Ester, M., Kriegel, H., J., S., & Xu, X. (1996, 1996 Aug 2). A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of KDD-96.

Focardi, S. M., & Fabozzi, F. J. (2001). Clustering economic and financial time series: Exploring the existence of stable correlation conditions. *Discussion Paper*. https://www.researchgate.net/profile/SergioFocardi/publication/2547390_Clustering_economic_and_financial_time_series_Exploring_the_existence_of_stable_correlation_conditions/links/56b379c608ae5deb2657d5ba/Clustering-economic-and-financial-time-series-Exploring-the-existence-of-stable-correlation-conditions.pdf

Forgy, E. W. (1965). Cluster analysis of multivariate data: Efficiency versus interpretability of classifications. *Biometrics*, *21*, 768-769.

Goldman Sachs. (2022). *Survey: Small business challenges worsen amid record inflation and workforce shortages*. Retrieved 30 Sept 2022 from https://www.goldmansachs.com/citizenship/10000-small-businesses/US/infographics/small-businesses-fear-looming-recession/index.html

González-Vidal, A., Barnaghi, P., & Skarmeta, A. F. (2018). BEATS: Blocks of Eigenvalues Algorithm for Time Series Segmentation. *IEEE Transactions on Knowledge and Data Engineering*, *30*(11), 2051-2064. https://doi.org/10.1109/TKDE.2018.2817229

Gopikrishnan, P., Plerou, V., Liu, Y., Amaral, L. A. N., Gabaix, X., & Stanley, H. E. (2000). Scaling and correlation in financial time series. *Physica A: Statistical Mechanics and its Applications*, *287*(3), 362-373. https://doi.org/https://doi.org/10.1016/S0378-4371(00)00375-7

Graves, D., & Pedrycz, W. (2010,. 2010). Proximity fuzzy clustering and its application to time series clustering and prediction. 2010 10th International Conference on Intelligent Systems Design and Applications.

Guha, S., Rastogi, R., & Shim, K. (1998). *CURE: An efficient clustering algorithm for large databases* Proceedings of the 1998 ACM SIGMOD international conference on Management of data, Seattle, Washington, USA. https://doi.org/10.1145/276304.276312

Guo, Y., Tang, D., Tang, W., Yang, S., Tang, Q., Feng, Y., & Zhang, F. (2022). Agricultural Price Prediction Based on Combined Forecasting Model under Spatial-Temporal Influencing Factors. *Sustainability*, *14*(17), 10483.

Guoqing, C., Qiang, W., & Hong, Z. (2001, 2001). Discovering similar time-series patterns with fuzzy clustering and DTW methods. *Proceedings Joint 9th IFSA World Congress and 20th NAFIPS International Conference* (Cat. No. 01TH8569).

He, W., Feng, G., Wu, Q., He, T., Wan, S., & Chou, J. (2012). A new method for abrupt dynamic change detection of correlated time series. *International Journal of Climatology*, *32*(10), 1604-1614. https://doi.org/https://doi.org/10.1002/joc.2367

Hotstats.(2022). *Hotels Are Filling Up. Guest Stomachs, Too.* https://www.hotstats.com/blog/hotels-are-filling-up-guest-stomachs-too

Ito, F., Hiroyasu, T., Miki, M., & Yokouchi, H. (2009, 20-24 Aug. 2009). Detection of preference shift timing using time-series clustering. 2009 IEEE International Conference on Fuzzy Systems.

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM Comput. Surv.*, *31*(3), 264–323. https://doi.org/10.1145/331499.331504

JP Morgan. (2022). *Midyear Survey: Business optimism falls to record lows*. https://www.jpmorgan.com/commercial-banking/insights/business-leaders-outlook-pulse-survey

Kalpakis, K., Gada, D., & Puttagunta, V. (2001, 29 Nov.-2 Dec. 2001). Distance measures for effective clustering of ARIMA time-series. Proceedings 2001 IEEE International Conference on Data Mining.

Karypis, G., Eui-Hong, H., & Kumar, V. (1999). Chameleon: Hierarchical clustering using dynamic modeling. *Computer*, *32*(8), 68-75. https://doi.org/10.1109/2.781637

Kaufman, L., & Rousseeuw, P. J. (2009). *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons Inc.

Klein, J. L. (1997). *Statistical visions in time: a history of time series analysis 1662-1938*. Cambridge University Press.

Kubatko, O., Merritt, R., Duane, S., & Piven, V. (2023). The impact of the COVID-19 pandemic on global food system resilience. *Mechanism of an Economic Regulation*, *1*(99), 144-148.

Kumar, R. P., & Nagabhushan, P. (2006). Time Series as a Point-A Novel Approach for Time Series Cluster Visualization. DMIN.

Lkhagva, B., Yu, S., & Kawagoe, K. (2006, 2006). New Time Series Data Representation ESAX for Financial Applications. 22nd International Conference on Data Engineering Workshops (ICDEW'06).

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability.

McAllister, J. (2022). *Operators face 'significant challenge managing costs and attracting footfall' as menu sizes and prices rise*. BigHospitality. from https://www.bighospitality.co.uk/Article/2022/05/20/Hospitality-operators-face-significant-challenge-managing-costs-and-attracting-footfall-as-menu-sizes-and-prices-rise

McLachlan, G. J., & Basford, K. E. (1988). Mixture models: Inference and applications to clustering (Vol. 38). New York: M. Dekker.

Murugesan, G., & Radha, B. (2023). An extrapolative model for price prediction of crops using hybrid ensemble learning techniques. *International Journal of Advanced Technology and Engineering Exploration*, *10*(98), 1.

Ng, R. T., & Han, J. (2014, 1994). Efficient and effective clustering methods for spatial data mining. Proceedings of the 20th International Conference on very Large Data Bases.

Oliveira, A. L., & Antunes, C. M. (2001). Temporal data mining: An overview. KDD Workshop on Temporal Data Mining.

Pavlidis, N. G., Plagianakos, V. P., Tasoulis, D. K., & Vrahatis, M. N. (2006). Financial forecasting through unsupervised clustering and neural networks. *Operational Research*, *6*(2), 103-127. https://doi.org/10.1007/BF02941227

Ramirez, L. (2022, 2022). Rising food prices taking a toll on Bay Area restaurant owners. *CBS News Bay Area*. https://www.cbsnews.com/sanfrancisco/news/rising-food-prices-taking-a-toll-on-restaurant-owners/

Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *26*(1), 43-49. https://doi.org/10.1109/TASSP.1978.1163055

Sarangi, P. K., Sinha, D., Sinha, S., & Mittal, N. (2021). Machine learning approach for the prediction of consumer food price index. In *2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)*. 1-6. IEEE.

Sfetsos, A., & Siriopoulos, C. (2004). Time series forecasting with a hybrid clustering scheme and pattern recognition. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *34*(3), 399-405. https://doi.org/10.1109/TSMCA.2003.822270

Sheikholeslami, G., Chatterjee, S., & Zhang, A. (1998). Wavecluster: A multi-resolution clustering approach for very large spatial databases. Proceedings of the 24th Conference on very Large Databases.

Siyou Fotso, V. S., Mephu Nguifo, E., & Vaslin, P. (2020). Frobenius correlation based u-shapelets discovery for time series clustering. *Pattern Recognition*, *103*, 107301. https://doi.org/https://doi.org/10.1016/j.patcog.2020.107301

Venkateswara Rao, K., Srilatha, D., Jagan Mohan Reddy, D., Desanamukula, V. S., & Kejela, M. L. (2022). Regression based price prediction of staple food materials using multivariate models. *Scientific Programming*, *2022*. https://doi.org/10.1155/2022/4572064

Wang, W., Yang, J., & Muntz, R. (1997). STING: A statistical information grid approach to spatial data mining. Proceedings of 23rd Conference on Very Large Databases.

Wang, X., Smith, K. A., Hyndman, R., & Alahakoon, D. (2004). *A Scalable Method for Time Series Clustering* (Technical Report Monash University).

Watson, N. J., Bowler, A. L., Rady, A., Fisher, O. J., Simeone, A., Escrig, J., & Adedeji, A. A. (2021). Intelligent sensors for sustainable food and drink manufacturing. *Frontiers in Sustainable Food Systems*, 5, 642786.

Wu, Y., Liu, S., Smith, K., & Wang, X. (2018). Using correlation between data from multiple monitoring sensors to detect bursts in water distribution systems. *Journal of Water Resources Planning and Management*, *144*(2), 04017084. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000870

Zhang, T., Ramakrishnan, R., & Livny, M. (1996). *BIRCH: An efficient data clustering method for very large databases* Proceedings of the 1996 ACM SIGMOD International Conference on Management of data, Montreal, Quebec, Canada. https://doi.org/10.1145/233269.233324