TIME SERIES CLUSTERING OF FERTILITY IN EUROPEAN UNION

Karolína Bakuncová – Tomáš Löster

Abstract

Time series clustering is a relatively new addition to clustering methods, allowing for the classification of objects based of the development of a chosen characteristic in time. The main goal of this paper is analysing fertility patterns in the EU using time series clustering and finding common traits amongst grouped up members. Additionally, the work briefly introduces the basics of time series clustering, from the pre-processing of data, calculation of similarity measures and the description of the most used measure and the evaluation of the resulting clustering results. The consensus for every analysed country is decreasing fertility rate; however, the shape of the decrease and subsequent developments vary. Total fertility rate of countries like Bulgaria or Czechia dropped later than in Austria or France, whilst countries such as Croatia and Greece or Cyprus and Malta experienced more gradual declines. In general, the fertility patterns were closely tied to the geographical positions of the clustered members, where for example countries in the west showed different traits in development from countries in the east or south. The countries also shared similarities in other socio-economic factors, such as GDP per capita, expenditure on social protection benefits of the evolution of female labour participation rate.

Key words: time series, clustering, fertility

JEL Code: C38, J13

Introduction

Cluster analysis has been and still is one of the most useful tools for finding new and at first glance hidden information about patterns and relationships in a dataset. The basic principles of cluster analysis can be summed up to several key points: choosing an appropriate distance measure that is compatible with the clustering algorithm, creating groups of observations where the members within a group share similar characteristics while being distinctly different from the rest of the groups. While this description lacks nuance when it comes to all the ins and outs of the process, it highlights the important steps in conducting cluster analysis.

With that said, the advancements in data analytics have introduced a need for methods that are able to deal with high dimensional data. This notion is complicated even further with the introduction of time series, with its dynamism and auto-correlation (Mori, Mendiburu, & Lozano, 2016).

Time series clustering is as such a natural step forward, where at its core is the idea that tries to combine the basic principles of cluster analysis with distance measures that are specifically created for a better representation of similarities between time series. Additionally, closely connected with time series clustering is the use of prototype functions as a practical tool for visually representing the main characteristics of the clustered time series.

For this purpose, popular programming languages, like R, have been adapting to this need, introducing packages designed for time series clustering. In this paper, all the practical applications have been done via the package **dtwclust**. The name of the package comes from one of the most widely used distance measures for time series clustering called Dynamic Time Warping (DTW). As such, most of the algorithms in this package are specifically tailored to the DTW distance (Sardá-Espinosa, 2023).

Subsequently, the main purpose of this paper is to present a brief overview of time series clustering analysis on a chosen dataset, whilst also showcasing some of the interesting conclusions unique to the chosen method of analysis. As for the dataset itself, it consists of 27 time series characterizing the developments in fertility of all the current the EU members from the year 1965 to 2021, sourced from the World Bank database.

1   Total fertility rate and its development

Most of the developing countries of today are currently facing the growing issue of declining total fertility rate, which is closely tied to several other factors, such as increasing female labour participation, depending on cultural context of the country its strength of traditional beliefs and marital rate, as well as other socioeconomical factors such as family support and childcare availability (Chesnais, 1998), or in the recent years the impact of COVID pandemic. As a result, by year 2021 the total fertility rate (TFR) of every single member of EU has dropped below the desired replacement rate of (1).

However, regardless of the cultural and historical specifics unique to the individual members of EU, it is safe to conclude that there are similar patterns in fertility development between the aforementioned countries throughout time. For example, certain countries experienced a much rapid decline since 1965 in TFR compared to other members, whose

decline shows a much steadier development. Others have gone through periods of decline and increase, while there are even examples of some countries experiencing a steady increase in the last years, albeit that increase has in most cases been overturned by COVID (**Fig.**1). These observations introduce the possibility of using time clustering to further analyse the similarities in development of fertility between the members of EU.

Fig. 1: Overview of fertility development for current EU members (1965 – 2021)
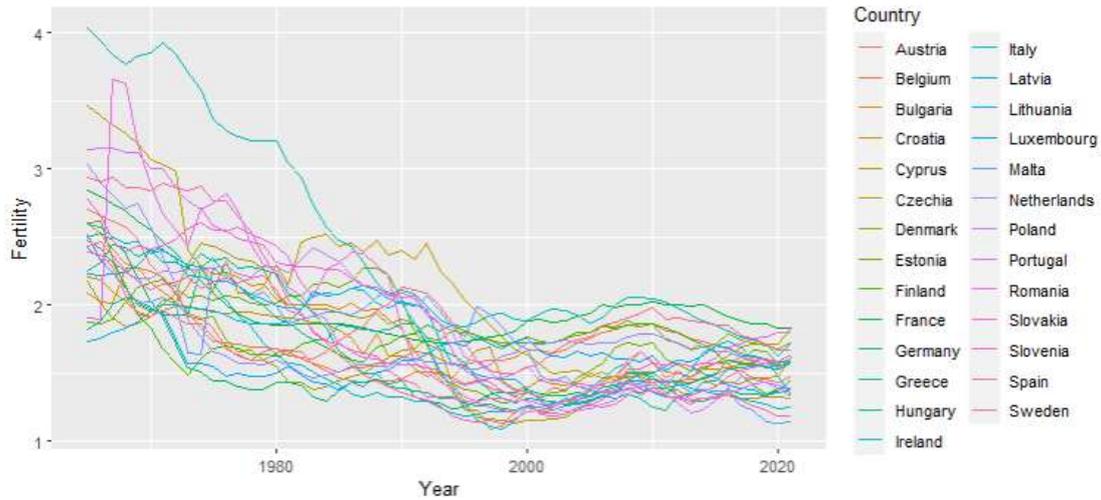
Source: World Data Bank, own calculations

2 Time series clustering

Unlike static data, the difficulty of working with dynamic data stems from its changes in time. Particularly, time series essentially provides information about an observed feature through a sequence of values sorted in a chronological order. Time series can also be described as a type of temporal data, which main characteristics include high dimensionality and large quantities of data. For simplicity's sake, in this paper the focus will be on univariate time series.

Overall, time series clustering can be defined as the process of unsupervised partitioning of a given dataset consisting of n time series data $D = \{T_1, T_2, \ldots, T_n\}$ into k homogenous subsets. Given that C_i, where $i = 1, 2, \ldots, k$, defines a cluster, it applies that $D = \bigcup_{i=1}^{k} C_i$ and $C_i \cap C_j = \emptyset$ for $i \neq j$ (Aghabozorgi, Wah, & Shirkhorshidi, 2015).

2.1 Clustering algorithms

In terms of novelty, the clustering algorithms used in time series clustering do not differ much from commonly used approaches in normal clustering analysis. Meaning, in time series

clustering, it is possible to categorize the algorithms methods into already established groups such as partitional or hierarchical clustering, which can be further divided into subgroups of agglomerative or divisive algorithms. As such, in most cases time series clustering adapts the commonly used clustering algorithms for static data and either uses a similarity measure appropriate for dynamic data or transforms the dynamic data in a way that results in static features (Liao, 2005). In this paper, the focus will be on showcasing hierarchical clustering.

While hierarchical clustering algorithms are limited in their capabilities, since it is difficult to control the final number of clusters, they have their benefits. For one, hierarchical clustering allows for clear visualization of created groups, as well as similarities between observed subjects. Other benefit of using hierarchical clustering stems from difficulty of obtaining labelled time series data. While partitional clustering offers more flexibility in group creation, in time series datasets one of the main issues is the missing information about the potential number of hidden groups. This is also one of the reasons why hierarchical clustering is commonly combined with another algorithm, to mitigate its flaws and build on its strengths. Additionally, hierarchical clustering, with the usage of appropriate similarity measures, allows for the clustering of time series of unequal length (Aghabozorgi, Wah, & Shirkhorshidi, 2015).

Nevertheless, the drawbacks or hierarchical clustering also include high computational complexity, so it is not advisable to use this algorithm on large time series.

2.2 Similarity measures

Measures of similarity, or dissimilarity, represent the main point of contention in time series clustering, as the definition of similarity for time series is the main tool for the application of already established clustering algorithms on time series. However, expressing similarity between time series comes with its own problems. Depending on whether time clustering is performed on raw data or its representation, there are different similarity measures to ensure compatibility. Moreover, unlike static data, the distance between time series is calculated approximately. Overall, it is possible to divide time series similarity measures into four groups, depending on the goal of the analysis and chosen procedure:

1) **Shape-based distances** focus on the similarity in the development of time series regarding the shape, rather than the question of when the development occurred (Esling & Agon, 2012). While similarity in time is a special case of similarity in shape, clustering based on shape matching is found to be more successful

(Ratanamahatana & Keogh, Three Myths about Dynamic Time Warping Data Mining, 2005). The most used metric in this category is Dynamic time Warping, based on an elastic transformation of time series allowing for better shape matching between time series (Senin, 2008). Another commonly used metric is Euclidian distance.

2) **Feature-based distances** rely on extracting certain features from raw data, such as Fourier or wavelet coefficients, autocorrelation values, etc. and comparing them.

3) **Structure-based distances** are based on either the similarity of fitted models for the time series in question or the amount of shared information between two time series. In both cases, the similarity measure is calculated on a basis of a model rather than initial series.

4) **Prediction-based distances** analyse the similarity of forecasts for the given time series (Mori, Mendiburu, & Lozano, 2016).

Dynamic time warping (DTW) is the preferred similarity measure in this paper. DTW strives to find an optimal path between two time series under given conditions by warping the points in between the starting and final positions via temporal distortion (Müller, 2007) (**Fig.**2). For this, DTW uses a distance matrix calculated for each individual points of two time series (**Fig.**3), where it tries to minimise the sum of distances in the matrix from the upper right corner to the bottom left corner.

Fig. 2: Comparison of Euclidian and DTW distances

Source: https://rtavenar.github.io/blog/dtw.html
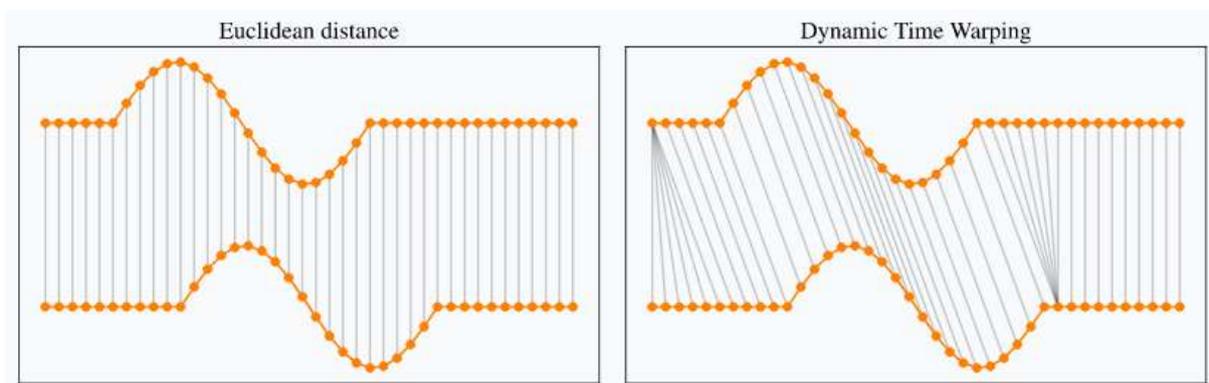
Fig. 3: DTW distance
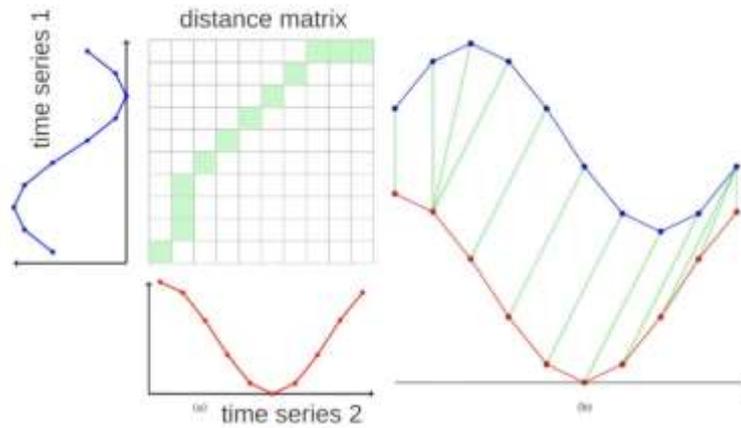
Source: https://youtu.be/_K1OsqCicBY

One of the pros of using DTW is the possibility of computing the distance between time series of varying length, albeit the results may prove to be subpar to interpolating time series to be equal in length in some situations (Ratanamahatana & Keogh, 2004). For more information about DTW and its implementation in R see (Giorgino, 2009).

2.3   Cluster prototypes

Especially in partitioning clustering algorithms, the quality of prototypes is essential for ensuring quality results (Aghabozorgi, Wah, & Shirkhorshidi, 2015). Cluster prototypes thus serve as cluster representatives, that can either be a part of the clustering procedure or be used for visual representation of a cluster average.

The computation of a cluster prototype is based on minimizing the distance between all the time series in a cluster with their prototype. The final cluster prototype is thus also a time series. Given that the cluster prototype is denoted as R_j, then a Steiner sequence is a time series that minimizes $E(C_i, R_j)$, where C_i represents a cluster and $F_1, F_2, ..., F_n$ are cluster members (Petitjean & Gançarsk, 2012) (Gusfield, 1998).

$$E(C_i, R_j) = \frac{1}{n}\sum_{x=1}^{n} dist(F_x, R_j), \qquad C_i = \{F_1, F_2, ..., F_n\} \qquad (1)$$

In general, there are a few ways how to acquire a prototype, either by assigning one of the cluster members (medoid as prototype), averaging the time series (average as prototype) or applying an iterative heuristic methods (local search prototype) (Aghabozorgi, Wah, & Shirkhorshidi, 2015) (Hautamaki, Nykanen, & Franti, 2008).

Since the distance measure used in this paper is DTW, as it is an elastic measure of similarity, simple averaging is unsuitable due to high computational complexity.

Nevertheless, there are several prototyping functions developed specifically for DTW distance, one of them being DTW barycenter averaging, shortened to DBA.

DBA is an iterative heuristic method. Simply put, the main goal of DBA is to minimize the sum of squared DTW distances in accordance with the definition of a Steiner sequence. DBA starts by assigning one of the series as a reference point, usually randomly. Each following iterations of DBA can then be divided into two parts:

1. Computing DTW between each individual series in a cluster and the temporary average sequence.
2. Refining the coordinates of the average sequence based on the barycenter of coordinates associated to it in the previous step.

It is important to note that during each iterations the associations between the average sequence and the coordinates of other series in a cluster may change. These changes are impossible to predict, thus why convergence is necessary. More details about this topic can be found in (Petitjean, Ketterlin, & Gançarski, 2011).

## 3   Practical applications

Next part of the paper will focus on the actual implementation of the previously discussed methods. Additionally, these methods will be applied to the same dataset twice, once on the raw data and again on transformed data, with the results compared afterwards, since in real practice, time series clustering is rarely performed on raw data.

The method of transformation used in this paper is knows as Z-normalization, or more precisely Normalization to Zero Mean and Unit of Energy. Normalization of a time series results in a vector with an approximate mean of zero and standard deviation of one and can be written as following:

$$x_i' = \frac{x_i - \mu}{\sigma} , \qquad i = 1,2,\ldots n \tag{2}$$

While transforming time series this way can lead to biased results, it allows for more accurate structural comparison of time series development, as instead of working with time series with wildly different means and standard deviations we convert them onto a single unified scale. On the other hand, that leads to its disadvantage in that it disregards the amplitude of the changes in time. More on that topic in (Kanellakis & Goldin, 1995). Visualization of Z-normalization can be seen in **Fig.** 4.

**Fig. 4: Effects of z-normalization of the time series**
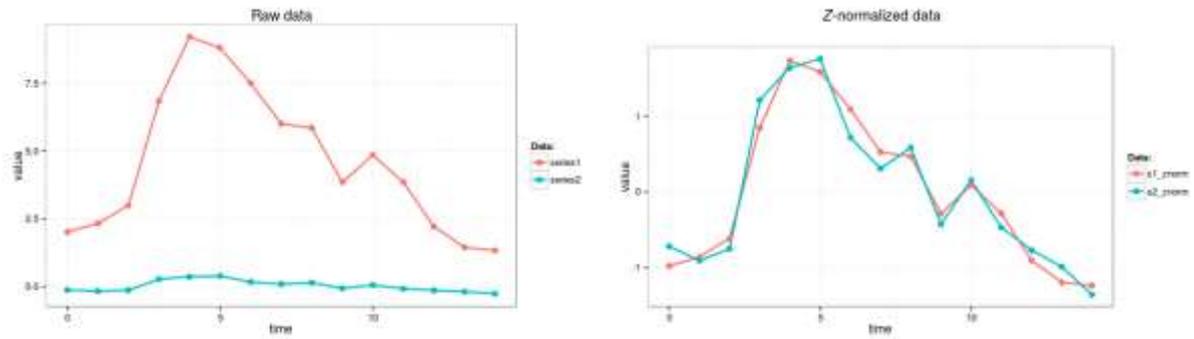
In both cases, time series clustering will be done using slightly modified Ward clustering where the differences are squared (Murphy, 2021), using DTW distance and DBA prototype function for graphical cluster representation.

2.1 Hierarchical clustering

Based on the dendrogram of time series clustering performed on the original dataset (**Fig.** 5), it is possible to differentiate several distinct groups of countries similar in their fertility, with one evident outlier being **Ireland**.

Fig. 5: Dendrogram (raw data)
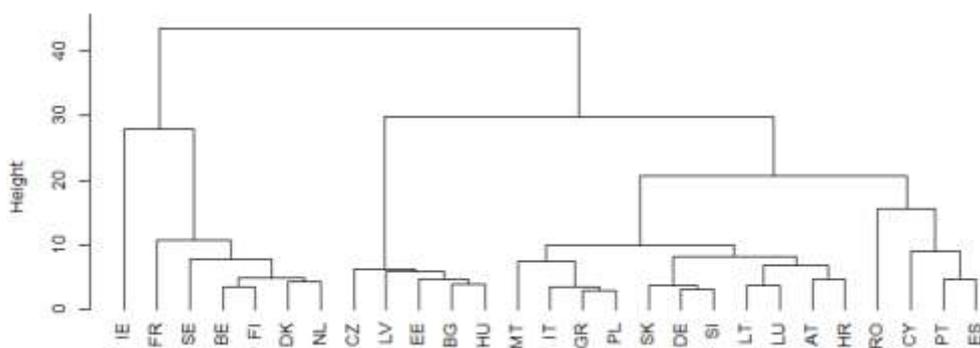
However, assuming we are more interested in the overall shape of the development, rather than the scale of changes fertility experiences over the observed period, we can use z-normalization. This allows for conversion of all the fertilities to a single scale (**Fig.** 6), which mostly eliminates any potentially strong outliers that could influence the quality of clustering. And while this introduces a certain bias, for some of the clustering algorithms transformation is a necessary step and, in most cases, does improve the overall results.

26

Fig. 6: Normalized fertility rates for all members of EU
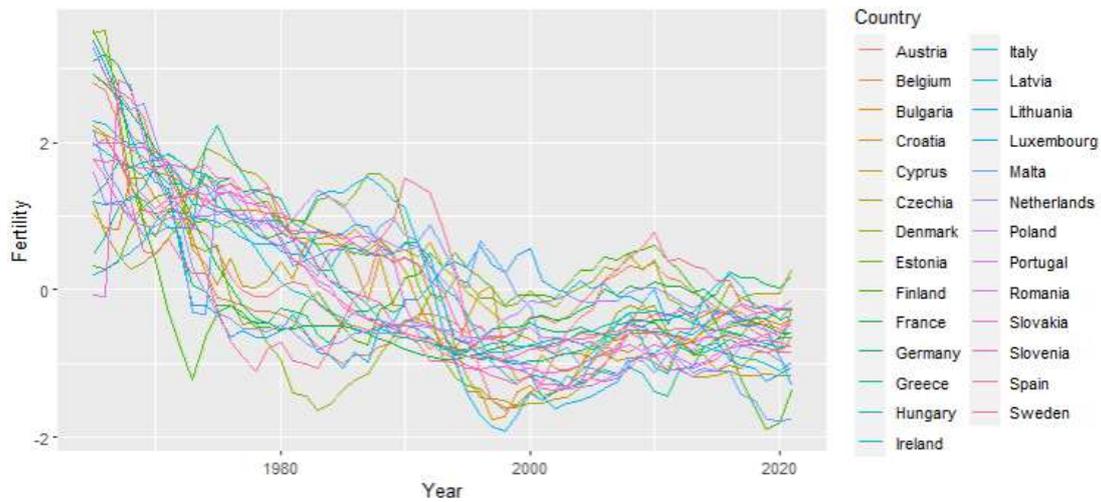
Source: World Data Bank, own calculations

There are other arguably more suitable options of similarity measures specifically for time clustering of normalized data than DTW, such as shape based distance (Paparrizos & Gravano, 2015), for the sake of simplicity, same clustering algorithm will be used on normalized data in this paper (**Fig.** 7).

Fig. 7: Dendrogram (normalized data)
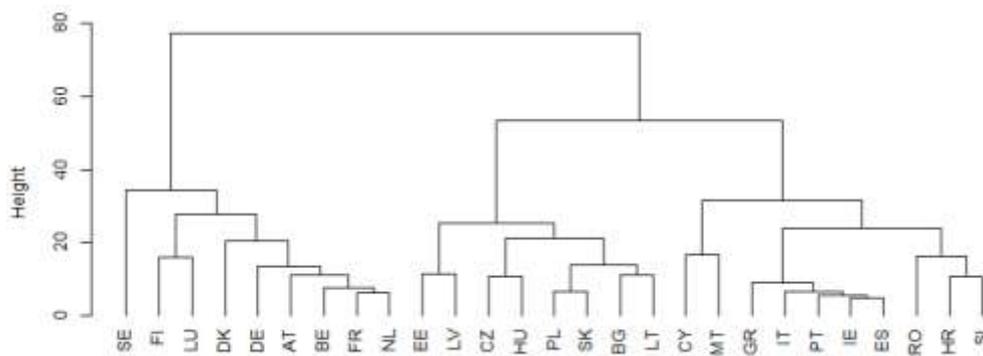
Looking at the final dendrograms for both raw and transformed data, while the final clusters differ from each other, we can still deduce some countries that are more likely to be clustered together, which speaks in favour of these countries sharing likely similarities in their fertility developments both in shape and magnitude of changing.

2.2   Comparison of clusters

Given the resulting dendrograms in **Fig.** 5 and **Fig.** 7, the optimal number of clusters is arguably similar in both cases, which is further supported when using evaluation measures

such as Silhouette score, Calinski-Harabasz index, Dunn index, Davies-Bouldin index and its modification or COP index. For the use in time series clustering in R, many of these commonly seen measures are already implemented in the dtwclust package.

Usually, as it is difficult to know which measure will work best with the selected time series clustering algorithm, several of them are used and the models are evaluated based on which model has the best measures overall (Sardá-Espinosa, 2023).

Based on the values for evaluation measures (**Tab**.1 and **Tab**. 2) data, when using Ward method of clustering, the optimal number of clusters in both cases is arguably not definite. However, six cluster in both cases seem to pose as a nice middle ground and will allow for a direct comparison.

Tab. 1: Evaluation measures using Ward method (raw data)

	2 clusters	3 clusters	4 clusters	5 clusters	6 clusters
Silhouette index (maximized)	**0.3553**	0.2715	0.2982	0.2990	0.3169
Calinski-Harabasz index (maximized)	**18.1030**	7.7919	6.8663	8.8467	7.6951
Davies-Bouldin index (minimized)	1.3242	1.4560	0.9785	0.9313	**0.7642**
Modified DB index (minimized)	1.3242	1.4911	0.9873	1.0325	**0.7776**
Dunn index (maximized)	0.1253	0.1253	0.1998	0.2456	**0.3154**
COP index (minimized)	0.4025	0.3307	0.2593	0.2031	**0.1712**

 * The best value of a given index in a row is marked in **bold**.

Tab. 2: Evaluation measures using Ward method (normalized data)

	2 clusters	3 clusters	4 clusters	5 clusters	6 clusters
Silhouette index (maximized)	**0.3370**	0.2612	0.2678	0.2265	0.2309
Calinski-Harabasz index (maximized)	14.2070	**14.4919**	10.2400	8.3834	7.5049
Davies-Bouldin index (minimized)	1.1667	1.2254	0.9826	1.0738	**0.9614**
Modified DB index (minimized)	1.1667	1.2673	**1.0042**	1.1735	1.0864
Dunn index (maximized)	0.2091	0.2619	0.2687	0.3071	**0.3752**
COP index (minimized)	0.4024	0.3345	0.3097	0.2780	**0.2575**

 * The best value of a given index in a row is marked in **bold**.

Based on the clustered time series and their given prototypes in **Fig.** 8 and **Fig.** 9, while the general developments in each cluster were not very clear for raw data, by using normalized data we are able to come to much defined and easily interpretable results.

Tab. 3: Countries based on their cluster assignment (ordered by raw data)

	AT	HR	DE	GR	IT	LT	LU	MT	PL	SK	SI
Raw	1	1	1	1	1	1	1	1	1	1	1
Normalized	1	3	1	3	3	2	5	4	2	2	3

	BE	DK	FI	FR	NL	SE	BG	CZ	EE	HU	LV
Raw	2	2	2	2	2	2	3	3	3	3	3
Normalized	1	1	5	1	1	6	2	2	2	2	2

	CY	PT	ES	IE	RO
Raw	4	4	4	5	6
Normalized	4	3	3	3	3

The first cluster has the most tangible difference in cluster assignment between raw and normalized data (**Tab**. 3). Before normalization, this cluster was also the most confusing and the least informative out of all the clusters, where it was difficult to ascertain what was the main development trend for the clustered members.

After normalization most of the previous cluster members have been classified into different groups, albeit some members were still grouped in the same group with each other (ex. Croatia, Greece, Italy, and Slovenia from the first cluster got linked with Portugal, Spain, Ireland, and Romania from other clusters).

In the case of the other clusters, second cluster members stayed mostly together (Belgium, Denmark, France, and Netherlands), although Finland and Sweden have been switched out by Austria and Germany.

Previous members of the third cluster even after normalization were classified together (Bulgaria, Czechia, Estonia, Hungary, and Lithuania) with additional countries from the previously first cluster (Latvia, Poland, and Slovakia).

As for the smallest clusters, unlike raw data, where two clusters had a single member, now there is only one such cluster formed by Sweden. In the case of raw data, these clusters were arguably not very informative when it comes to the comparison of shapes in fertility developments, as there were no evident differences between those two countries and other quite similar clusters, other than the fact that these two countries had a much higher fertility at the beginning.

Nevertheless, this issue is remedied by scaling all the time series to a single scale, where now the structural differences are much clearer. Why Sweden was classified into a separate group from the others can be explained by Sweden experiencing a big jump in fertility in the 1990s, unlike other countries of the EU, which do not show any similar developments, as well as having other cultural and historical characteristics unique to the

region. All this led to visible improvements for the last few clusters. Before, Cyprus in cluster number four clearly showed different tendency of development, unlike Portugal and Spain (**Fig.** 8). Now, Portugal and Spain have been reclassified into a much more fitting cluster number three, and Cyprus was instead grouped with Malta (**Fig.** 9).

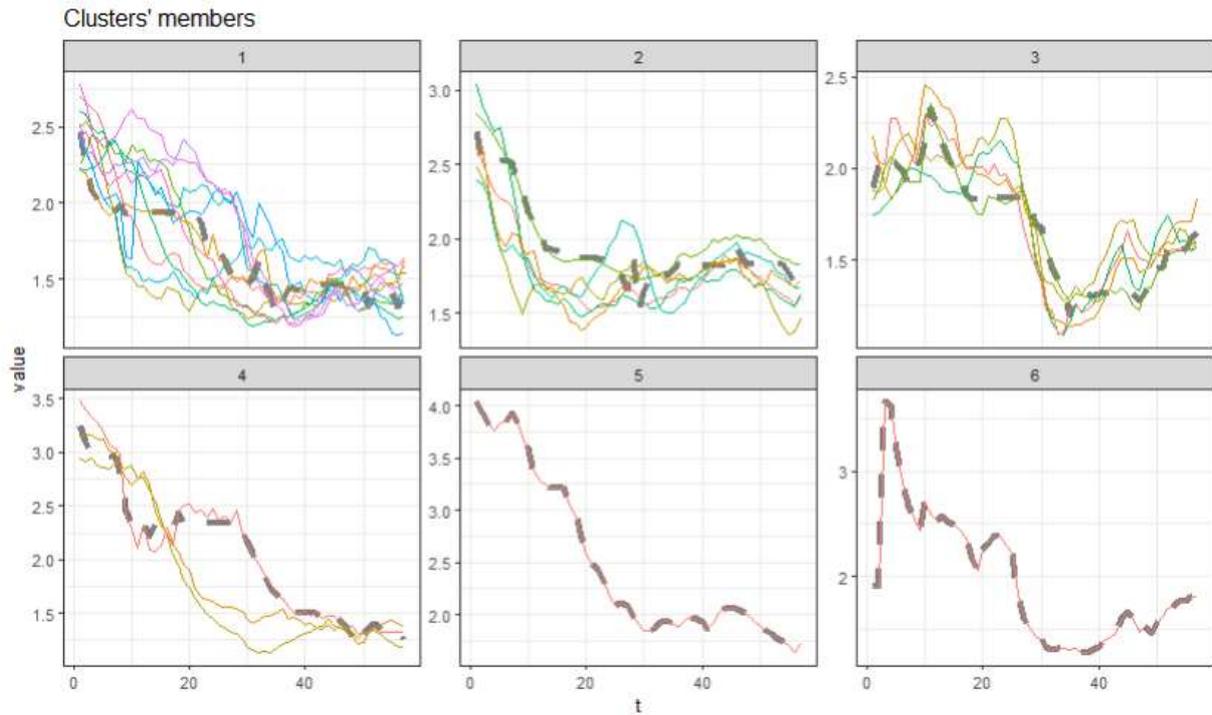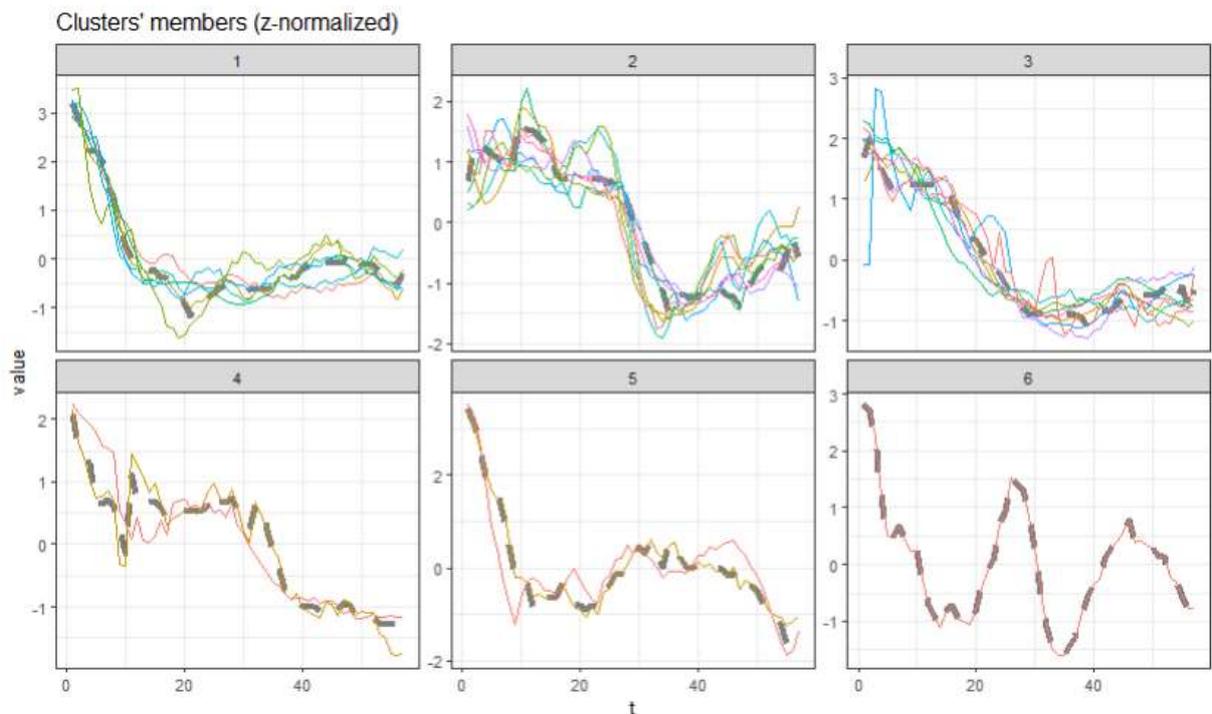**Fig. 8: Six clusters and their prototype functions (raw data)**

**Fig. 9: Clusters and their prototype functions (normalized data)**

2.3   Cluster characteristics

Looking at each cluster more closely, it is possible to conclude, that the distinctive shapes of fertility development often is the result of similar circumstances, be it shared geographical traits, history, societal customs, or economic background.

First such cluster is made up countries in the western Europe: Austria, Belgium, Denmark, France, Germany, and Netherlands (**Fig.** 10). Fertility development is characterized by rapid drop until 1985, with either stagnant or slightly increasing development afterwards. Moreover, countries of western Europe are in general quite wealthy with high GDP per capita and have good social protection systems in place. The drop in fertility is also followed by a growing percentage of working women, where in 2021 around 60 % to 80 % of all women in the country aged 15 to 64 years participated in labour.

The second cluster (**Fig.** 10) is made up mostly of countries in central and eastern Europe: Bulgaria, Czechia, Estonia, Hungary, Latvia, Lithuania, Poland, and Slovakia. For these countries, the rapid drop in fertility happened later and continued throughout the 1990s. However, by early 2000s the fertility started to rise and continues to do so until today. Additionally, unlike the previous cluster, members of this cluster are less wealthy and do not have as much funding for their social benefits. On the other hand, the development in female labour participation rate is similar (65 – 80% by 2021).

Fig. 10: First and second cluster
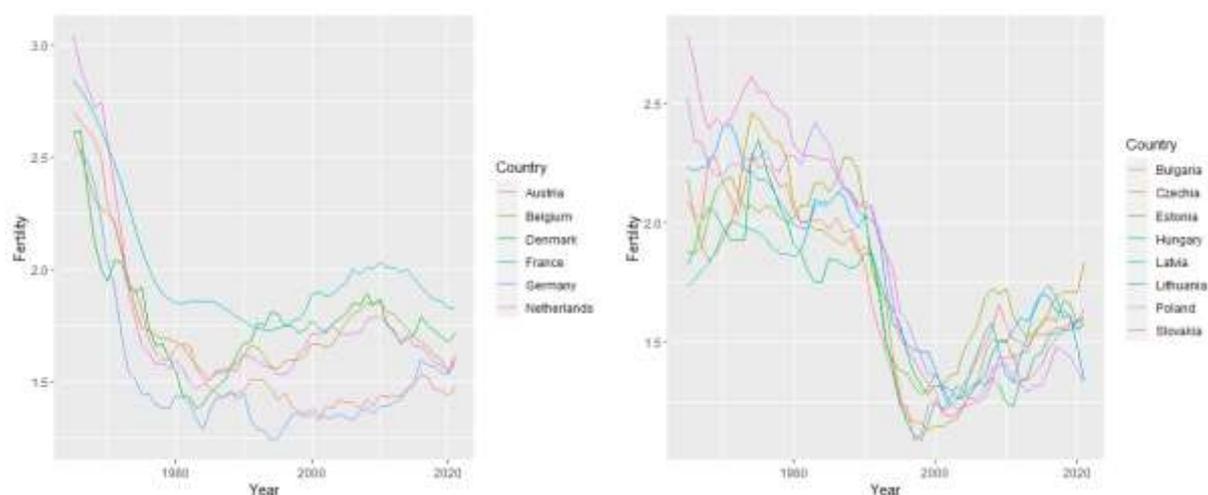
Source: World Data Bank, own calculations

The third cluster (**Fig.** 11) is mostly made up of countries in the south: Croatia, Greece, Italy, Portugal, Romania, Slovenia, and Spain, with the exception of Ireland. Unlike the previous two clusters, the development of fertility in the third cluster does not feature

any steep drops. Contrarily, the development has a much more gradual decline, that stops at the beginning of 2000s. However, it does not start to rise again significantly, keeping fertility firmly bellow the replacement rate. Except Italy and Ireland, countries in this cluster are also not that wealthy and do not have high expenditures on social protection benefits. As for female labour participation rate, it is the lowest of all the clusters, around 55 % to 65 % by 2021.

When it comes to the fourth cluster (**Fig.** 11), it presents a unique case where both of its members are island countries in the Mediterranean sea. It would be safe to assume that the similarity in the development of fertility in this case has to do with the geographical and historical circumstances, rather than its current socio-economical situation. Nevertheless, these countries have high female labour participation rate (around 75 % in 2021), mediocre funding of social protection benefits and not a lot of overall wealth.

**Fig. 11: Third and fourth cluster**
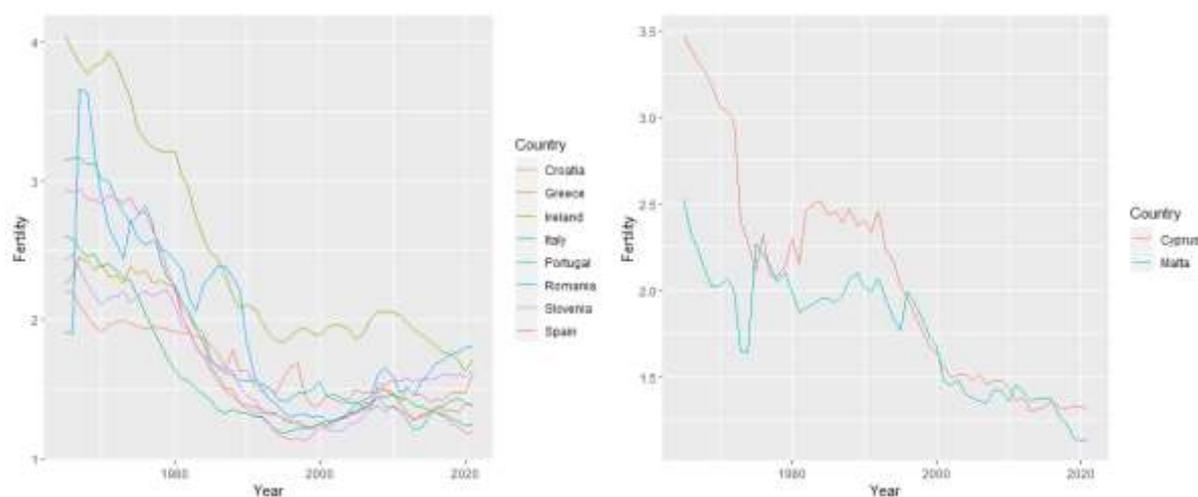
Source: World Data Bank, own calculations

While being similar, the fifth cluster (**Fig.** 12) also having an early drop in fertility comparable to the first cluster, the drop is noticeably faster, as well as being followed by a more distinct increase and subsequent decrease by 2020. Luxembourg and Finland are amongst the wealthiest in the Eruopean Union, additionally having high expenditures on social protection benefits. Female labour participation rate here is also quite high, around 70 % to 80 % in 2021.

The last cluster features an entirely unique case of fertility development, not shared by any other member of the EU (**Fig.** 12). Sweden features several rapid drops and increases in its fertility throughout the observed period of time. This development mirrors the

development of female labour participation rate, where high fertility meant high female labour participation rate, which is very unusual. As such, it would make sense to put Sweden in an entirely separate group.

Fig. 12: Fifth and sixth cluster
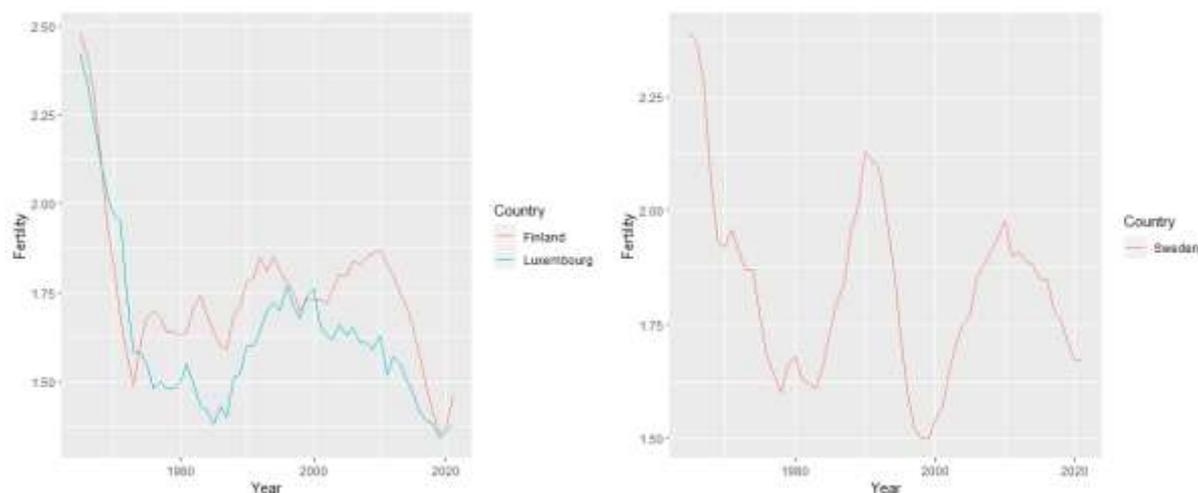
Source: World Data Bank, own calculations

Conclusion

Time series clustering is a useful tool when trying to ascertain the presence of distinct patterns of development in data. At its core, the basic clustering procedure is still the same. However, since instead of working with single points the analysis is applied to entire sequences of values ordered in time, certain modifications are needed, mainly to the calculations of similarity between two time series and the cluster average.

Clustering generally yields better results if applied on transformed data, since time series commonly feature components that may negatively affect the clustering outcome, be it noise, difference in scale, etc. One of these transformations is normalization.

The most common measure of similarity used in time series clustering is Dynamic Time Warping (DTW), which allows for much more accurate definition of similarity when it comes to the comparison of time series shapes. Moreover, this distance measure is generally compatible with all sorts of clustering algorithms, be it hierarchical or partitional, albeit at certain times the use of other specific distance measures may give better results.

When it comes to the clustering algorithm and evaluation measures, steps and methods already well established in normal clustering may be applied in time series clustering as well, if the dataset does not call for any specific procedures.

By applying time series clustering on a dataset featuring total fertility rates for all the current members of the EU, it was possible to create six distinct groups of development patterns. Each cluster has its own set of common characteristics. Most clusters feature distinct geographical similarities, where each cluster can be defined to be made up of mostly western, eastern, or southern countries, with additional exceptions. Groups also commonly share similarities in socio-economic backgrounds.

In general, the clustering may be labelled as successful, although more detailed examination of characteristics for each group is needed to come to more conclusive results about shared fertility development patterns.

References

Aghabozorgi, S., Wah, T., & Shirkhorshidi, A. S. (2015, May). Time series clustering – A decade review. *Information Systems*(53), 16-39.

Chesnais, J. C. (1998). Below-Replacement Fertility in the European Union (EU-15): Facts and Policies, 1960 - 1997. *Review of Population and Social Policy*(7), 83-101.

Esling, P., & Agon, C. (2012, November). Time series Data Mining. *ACM Computing Surveys, 45*(1).

Giorgino, T. (2009, August). Computing and Visualizing Dynamic Time Warping. *Journal of Statistical Software, 31*(7), 1-24.

Gusfield, D. (1998). *Algorithms on Stings, Trees, and Sequences: Computer Science and Computational Biology.* Cambridge University Press.

Hautamaki, V., Nykanen, P., & Franti, P. (2008). Time series Clustering by Approximate Prototypes. *2008 19th International Conference on Pattern Recognition* (pp. 1-4). Tampa, FL, USA: IEEE.

Kanellakis, P. C., & Goldin, D. Q. (1995). On Similarity Queries for Time series Data: Constraint Sp eci cation and Implementation. *International conference on principles and practice of constraint programming* (pp. 137-153). Berlin, Heidelberg: Springer Berlin Heidelberg.

Liao, T. W. (2005, November). Clustering of time series data—a survey. *Pattern Recognition, 38*(11), pp. 1857-1874.

Mori, U., Mendiburu, A., & Lozano, J. A. (2016, December). Distance Measures for Time Series in R: The TSdist Package. *The R Journal, 8*(2).

Müller, M. (2007). Dynamic Time Warping. In M. Müller, *Information Retrieval for Music and Motion* (pp. 69-84). Berlin, Heidelberg: Springer.

Murphy, P. (2021, April 25). *Clustering Data in R.* Retrieved from RPubs: https://rstudio-pubs-static.s3.amazonaws.com/599072_93cf94954aa64fc7a4b99ca524e5371c.html

Paparrizos, J., & Gravano, L. (2015). k-shape: Efficient and accurate clustering of time series. *Proceedings of the 2015 ACM SIGMOD international conference on management of data*, (pp. 1855-1870).

Petitjean, F., & Gançarsk, P. (2012). Summarizing a set of time series by averaging: From Steiner sequence to compact multiple alignment. *Theoretical Computer Science*, pp. 76-91.

Petitjean, F., Ketterlin, A., & Gançarski, P. (2011). A global averaging method for dynamic time warping, with applications to clustering. *Pattern recognition, 44*(3), pp. 678-693.

Ratanamahatana, C. A., & Keogh, E. (2004). Everything you know about dynamic time warping is wrong. *Third Workshop on Mining Temporal and Sequential Data*, (pp. 1-11). Seattle, WA.

Ratanamahatana, C. A., & Keogh, E. (2005, April). Three Myths about Dynamic Time Warping Data Mining. *Proceedings of the 2005 SIAM International Conference on Data Mining.*

Sardá-Espinosa, A. (2023). *dtwclust: Time Series Clustering Along with Optimizations for the Dynamic Time Warping Distance.* Retrieved from CRAN: https://CRAN.R-project.org/package=dtwclust

Senin, P. (2008). *Dynamic Time Warping Algorithm Review.* University of Hawaii at Manoa, Information and Computer Science Departmen, Honolulu, USA.

Contact

Bakuncová Karolína, Bc.

Prague University of Economics and Business

W. Churchill Sq. 1938/4

130 67 Praha 3 – Žizkov, Czech Republic

bakk02@vse.cz

Löster Tomáš, Ing., Ph.D

Prague University of Economics and Business

W. Churchill Sq. 1938/4

130 67 Praha 3 – Žizkov, Czech Republic

tomas.loster@vse.cz