

Augustyński, Iwo; Laskoś-Grabowski, Paweł

Preprint

Clustering Macroeconomic Time Series

Suggested Citation: Augustyński, Iwo; Laskoś-Grabowski, Paweł (2017) : Clustering Macroeconomic Time Series, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel und Hamburg

This Version is available at:

<https://hdl.handle.net/10419/171380>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Clustering Macroeconomic Time Series

Iwo Augustyński* Paweł Laskoś-Grabowski†

November 27, 2017

Abstract

There is growing literature in macroeconomics, especially on business cycle synchronization, employing different methods of time series clustering. However, even as an unsupervised learning method, this technique requires making choices that are nontrivially influenced by the nature of the data involved. By extensively testing various possibilities, we arrive at a choice of a dissimilarity measure (compression-based dissimilarity measure, or CDM) which is particularly suitable for clustering macroeconomic variables. We check that the results are stable in time and consistent with the literature on core-periphery pattern of European business cycles. We also successfully apply our findings to the analysis of national economies, specifically to identifying their structural relations. To our knowledge, it is the first comprehensive analysis of the usefulness of the different dissimilarity measures for the macroeconomic research.

JEL: E00, C18, C63

Keywords: time series clustering, similarity, cluster analysis, GDP

1 Introduction

The algorithms for clustering similar time series or, more generally, similar high-dimensional sequences, are important in areas as diverse as biomedicine, computational biology, electronic manufacturing, physics, seismology and speech recognition.

Econometrics could also benefit from a vast research effort made in these and other areas. For example, according to Focardi and Fabozzi (2004), clustering of economic and financial time series includes the following areas of application:

- identifying areas or sectors for policy-making purposes;
- identifying structural similarities in economic processes for economic forecasting;
- identifying stable dependencies for risk management and investment management.

*Wrocław University of Economics, Wrocław, Poland
e-mail: iwo.augustynski@ue.wroc.pl

†Institute of Theoretical Physics, University of Wrocław, Poland

For studies in macroeconomics and economics policy, one of the most promising advantages of time series clustering is its ability to identify structural similarities in the processes that generate time series at different points in time and space. This method also allows for presentation of results in the form of easy-to-understand dendrograms.

The aim of this paper is to find the most appropriate dissimilarity measure for macroeconomic clustering analysis.

The problem of time series similarity could be boiled down to measurement of the comovement of macroeconomic aggregates. The most popular approaches to this issue are (Croux, Forni, & Reichlin, 2001; Haan, Inklaar, & Jong-A-Pin, 2008):

- correlation;
- cointegration, that is, the existence of a linear combination of the two processes that is stationary;
- codependence, which refers to linear combinations of correlated processes that are of lower autoregressive order than others;
- common features, that is, linear combinations that are unpredictable with respect to past information, and common cycles which are defined as common features in first differences for processes that are cointegrated.

According to Croux et al. (2001), these concepts pose several problems. First, correlations might be detected where no correlation is present. Second, high cross-correlation neither implies nor is implied by cointegration, common cycles, or common features. Third, these three measures are binary. For example, two processes are either cointegrated or not, but different degrees of association cannot be established. Tests such as the Johansen test, performed by commercial econometric packages, consist of fitting empirical data to cointegrated models such as Error Correcting Models (ECM). In the case of a large number of time series, cointegration is a rather cumbersome exploratory methodology.

The contribution of the present paper to the literature is as follows:

- to our knowledge, it is the first comprehensive analysis of the usefulness of the different dissimilarity measures for the macroeconomic research,
- it offers ready-to-use methodology,
- it offers tools to present data in easy-to-understand dendrograms,
- it is provided with code and web application for easy use¹.

The remainder of this paper is organized as follows: in Section 2 we evaluate available dissimilarity measures and propose CDM (compression-based dissimilarity measure) as a solution for clustering of the macroeconomic time series. In Section 3 we check the robustness of the presented methodology by applying the proposed clustering method and comparing created clusters with the literature. Finally, in Section 4 we present concluding remarks.

All figures are the results of own calculations based on R package TSclust and Eurostat data (namq_10_gdp dataset).

¹<http://iwoaugustynski.ue.wroc.pl/apps/TSclustering/>

2 Experimental Evaluation of Dissimilarity Measures

A well-known data mining technique, clustering is an example of *unsupervised* learning: a clustering algorithm creates clusters as a function of its internal rules (whereas in supervised learning the algorithm learns from “known” examples). The objective of clustering is to create groups of objects that are close to each other and distant from other groups of objects. If distance corresponds to similarity, clustering forms groups of objects that are maximally similar.

2.1 Design of the Experimental Approach

A dissimilarity measure suitable for our applications could be informally described as attaining low values for pairs of time series that exhibit a causal relationship. This property is hard, if not outright impossible, to formulate in rigorous terms, but it can be approximated well enough by demanding that the measure is insensitive to *translating* the series in time. Importantly, it *should* be sensitive to other time transforms, including warping (acceleration), whether uniform or not. Conversely, other kinds of transforms, such as scaling the values by a constant, adding a constant to the values, or adding noise to the values, should not, in principle, affect the measure much.

We can apply this observation to design an experiment to identify prospective measures. For a given time series, we can compute the distance separating it from its delayed copy, as well as the distance separating it from its warped copy. The ratio of these distances is a measure of how well a dissimilarity measure performs. We are looking for a measure that performs well for as many as possible time series, delays, and warp factors. Furthermore, this should hold true even if the delayed copy of the series is additionally subjected to *perturbations* such as scaling, shifting, noise, or a combination thereof.

To formalize the idea, we fix a time series $X = \{X_i\}_{i=1}^T$, dissimilarity measure M , delay $\delta \in \mathbb{N}$, and warp factor $\alpha \in \mathbb{R}$. We call the subseries $B[X; \alpha] = \{X_i\}_{i=1}^{\lfloor T/\alpha \rfloor}$ and $D[X; \alpha, \delta] = \{X_i\}_{i=1+\delta}^{\lfloor T/\alpha \rfloor + \delta}$ respectively the *base* and *delayed* series. We also obtain from X the *warped* series $W[X; \alpha] = \{\bar{X}_i^{(\alpha)}\}_{i=1}^{\lfloor T/\alpha \rfloor}$ by taking averages of every α consecutive elements, trivially extending the notion for noninteger α . Strictly speaking, $\bar{X}_i^{(\alpha)} = \frac{1}{\alpha} \sum_{j=1}^T w_{ij}^{(\alpha)} X_j$, where

$$w_{ij}^{(\alpha)} = \begin{cases} \alpha & \text{if } j-1 < (i-1)\alpha < i\alpha < j \\ j - (i-1)\alpha & \text{if } j-1 < (i-1)\alpha < j < i\alpha \\ 1 & \text{if } (i-1)\alpha < j-1 < j < i\alpha \\ i\alpha - (j-1) & \text{if } (i-1)\alpha < j-1 < i\alpha < j \\ 0 & \text{otherwise.} \end{cases}$$

Let $d_M(\cdot, \cdot)$ be the distance between two time series under the measure M . We will use the ratio $R(M, X, \delta, \alpha) = \frac{d_M(B[X; \alpha], D[X; \alpha, \delta])}{d_M(B[X; \alpha], W[X; \alpha])}$ as a quality indicator, where $\tilde{D}[X; \alpha, \delta]$ is $D[X; \alpha, \delta]$ with perturbation applied. Note that X is truncated to its prefix $B[X; \alpha]$ for calculating R to fulfill a requirement of many of the measures that the series compared be of the same length.

Note that while the values of distances under different dissimilarity measures are not directly comparable to one another (as different measures may e.g. attain values in different ranges), ratios such as R are. Note also that if a measure is *good*, that is, in general it separates warped series more than the delayed series, its values of R will be

in general (or, ideally, always) less than 1. Note finally that a dissimilarity measure is *better* than another if its values of R are in general smaller.

2.2 Sampled Measures, Series, Parameters, and Perturbations

In our computations, we consider all dissimilarity measures provided by the R package TSdist, which, for the sake of simplicity and avoiding potential cognitive bias, may be calculated without supplying extra parameters. This is possible either thanks to their absence, or default values, or heuristics. These 24 measures are (referred to by names used within TSdist): `euclidean`, `manhattan`, `infnorm`, `ccor`, `sts`, `dtw`, `fourier`, `acf`, `pacf`, `ar.lpc.ceps`, `ar.mah.statistic`, `ar.pic`, `cdm`, `cid`, `cor`, `cort`, `wav`, `int.per`, `per`, `ncd`, `spec.glk`, `spec.isd`, `spec.llr`, and `pd`.

Furthermore, we consider warp factors α between 1.4 and 3.0 inclusive in steps of 0.2, and delays Δ of 2 through 10 quarters. Note that for series with quarterly data, the delay in terms of data points is $\delta = \Delta$, while for monthly data it is $\delta = 3\Delta$.

We also consider *separately* two different sets of time series, along with perturbations specific for each set. The first set contains 52 time series of *absolute-valued* data:

- Eurostat quarterly GDP values for 28 EU Member States,
- Eurostat quarterly GDP component values for 11 components of the UK GDP (with total GDP “component” omitted, as it is present in the preceding category),
- three series obtained from the FRED quarterly UK GDP values by reversing the sign as well as concatenating prefixes and suffixes of this and the original series,
- FRED monthly long-term (ten-year) government bond yields for USA, Germany, and France,
- FRED monthly short-term (three-month) certificates of deposit yields for USA and interbank rates for Germany and France,
- four artificial series: sine and triangular waves of three periods (each of 100 “months” for the sake of compatibility), with either constant or linearly diverging extrema.

The following perturbations were applied to the delayed copies of the series from this set in distinct computation runs:

- multiplying the values by a constant,
- adding a constant (proportional to the standard deviation of the original series) to the values,
- adding random noise (also proportional to the standard deviation) to the values,
- all of the above,
- none of the above.

The second set contains 45 time series of *quarter-to-quarter percentage changes* of the same quantities as in the first set, except the three artificial modifications of the UK GDP, as well as four of the UK GDP components for which the percentage data

is not available (i.e. compensation of employees, taxes on production and imports less subsidies, changes in inventories and acquisitions less disposals of valuables, and operating surplus and mixed income, gross). Also, the four artificial (sine and triangular wave) series are not converted to percentage changes, but (with different value ranges) treated as percentage change series in their own right. Given the different nature of the data in this set, we deem the perturbations of scaling and shifting inapplicable here, and computations are performed in two runs (without perturbations and with random noise, not adjusted for standard deviation in this case).

In each of the seven runs, ratios are computed for all possible combinations of measures, warp factors, delays, and time series.

2.3 Evaluation Results

Examining the results of all seven computation runs, we have primarily focused on the following quantities:

- $\max_{X,\Delta,\alpha} R$ for every measure M , that is, the maximum value of ratio R achieved across all time series, delays, and warp factors;
- the count of how often (out of 81 possible combinations of delay Δ and warp factor α) does measure M rank either first or in top five (of 24 measures considered) when ordered by $\max_X R$ (for given M, Δ, α).

Based on this, we arrive at the conclusion that `cdm` is the overall best performing measure. Specifically (see also Table 1):

- Without perturbations, both for absolute- and percentage-valued series, `cdm` is the only measure except for `ncd` that never exceeds 1 (and thus is *good* in the sense outlined above). Note that in these runs `ncd` performs better than `cdm`, with lower global maximum and more frequent appearances in the top spot or top five spots. In fact, `cdm` never ranks first in these cases, although only three measures ever do (`ncd`, `dtw`, and, only once for absolute- and five out of 81 times for percentage-valued series, `pdv`), and ranks third most often in the top five spots (behind `ncd` and `dtw` again).
- In the remaining five runs, `cdm` has the lowest global maximum, which additionally is always less than ca. $\frac{4}{3}$. For two runs (absolute-valued series with scaling and with all perturbations at once), it is the only measure with a maximum less than 2, while additionally for absolute-valued series with shifting `ncd` is the only other measure with a maximum less than 2.
- For absolute-valued series with all perturbations at once, `cdm` ranks most often in the top spot, and, importantly, it does so for more than half of the possible combinations of warp factor and delay. Additionally, it ranks second most often in the top spot for absolute-valued series with scaling and with shifting (behind `per`), and for absolute- and percentage-valued series with noise (behind `dtw`).
- For four runs (absolute-valued series with scaling, with shifting, and with all perturbations at once, as well as percentage-valued series with noise), `cdm` ranks most often in top five spots (although on par with `pdv` for absolute-valued series with shifting). For absolute-valued series with scaling and with all perturbations at once, it always ranks in top five spots, while for percentage-valued series with

noise, it fails to do so for only three combinations of warp factor and delay. For absolute-valued series with noise, `cdm` ranks third most often in top five spots (behind `dtw` and `cid`).

Another criterion that could be sensibly used here is the count of how often the ratios for a given measure exceed 1. Admittedly, with perturbations present, `cdm` does not perform well in this aspect. However, precisely in the presence of perturbations this requirement can be argued to be too restrictive: the delayed *and* perturbed series can be informally considered as at a disadvantage compared to its warped and *not* perturbed counterpart (especially if perturbations are, in some sense, large). It also does not give consistent conclusions across different perturbations: `pd` ranks best for absolute-valued series with scaling or shifting, `dtw` for absolute- and percentage-valued series with noise, and `int.per` for absolute-valued series with all perturbations at once. In general, for absolute-valued series, measures performing well with noise mostly perform poorly with other two perturbations, and vice versa. Finally, measures should perform well also without perturbations, and that in these cases `cdm` together with `ncd` rank best as the only *good* measures.

CDM (compression-based dissimilarity measure) introduced by Keogh, Lonardi, and Ratanamahatana (2004) and further elaborated upon in Keogh et al. (2007), has already been demonstrated (e.g. in these two references) to be immensely useful in various data mining aspects, prominently including time series clustering. It warrants mention, however, that it is not a *distance measure* (which is why we avoid using that phrase altogether throughout this paper), as, among other properties, by definition it attains values in the range $[\frac{1}{2}, 1]$, i.e. does not reach 0, even for equal arguments. This also means that the ratios R we compute in our experiment are always going to be within the range $[\frac{1}{2}, 2]$ for CDM, and it may be argued that the setup is biased in favour of this measure. However, we uphold our conclusion because no other measure emerges as a clear alternative, especially if we demand that it performs well both with noise and with other types of perturbations. As a bonus, CDM also performs well with both the absolute- and percentage-valued series, making this measure much more versatile, even though we do not explicitly require such behaviour. Also, the unambiguously good results without perturbations (i.e. ratios below 1) and aforementioned reports of CDM suitability for clustering in general support this choice.

2.4 Choice of Clustering Method

The next step in the procedure was to choose the most suitable clustering method. After clustering one should obtain a figure resembling a tree with some main branches and many smaller side branches. Conversely, a structure of ascending steps would disqualify a given clustering method.

We tested the following approaches:

- single linkage,
- complete linkage,
- Ward,
- average (UPGMA),
- McQuitty (WPGMA),

- median (WPGMC),
- centroid (UPGMC).

Only the first five of these produced useful dendrograms. Ultimately we decided to pick the Ward method, as it is less sensitive to changes in the length of the time series and creates better separated groups.

3 Empirical Application

For the purposes of this paper we employed the time series of GDP for analysis of the measure's stability, and its direct components (GDP, investment, consumption, import, export, employment, wages, etc.) to compare the economic structures of four EU Member States: Germany, France, Italy, and Spain. The reasoning for this choice is as follows: all of them are members of the Eurozone, Germany and France are the most advanced European countries, whereas Italy and Spain are the biggest lagging-behind members of the EU.

In order to establish international comparisons, we used data from the Eurostat after the introduction of the euro currency, that is, from the first quarter of 2000.

We decided to use the "raw" data, that is, quarterly, not seasonally adjusted time series. We argue that using data subject to preprocessing such as seasonal adjusting or detrending could introduce artificial distortions to the time series (Haan et al., 2008; Hamilton, 2017).

3.1 Type of Data

In the following calculations we used nominal GDP in millions of euro, but other options are also available:

- real (chain linked) values,
- national currency,
- first differences (or percentage change).

Which one is the most suitable for the considered method?

First, it depends on data availability. Usually the most up-to-date and the most accessible data is represented in nominal values in a given national currency.

Second, when comparing the behaviour of different real variables we are usually not interested in their nominal values, but their changes. Therefore, percentage change is the one most commonly used. But it is problematic in the case of, for instance, financial data, which is often presented as a percent rate (interest rates, yields, etc.). Percentage change of a percent rate may be hard to interpret.

Third, the answer to the question if real or nominal values are preferable is much more straightforward. If the dataset contains variables measured both in real terms (such as the number of unemployed) and in currency terms (such as consumption or GDP), chain-linked values should be employed. Otherwise, when all variables are represented in currency terms, distortions caused by inflation could be probably neglected.

We argue that the proposed dissimilarity measure is suitable in all of the aforementioned cases.

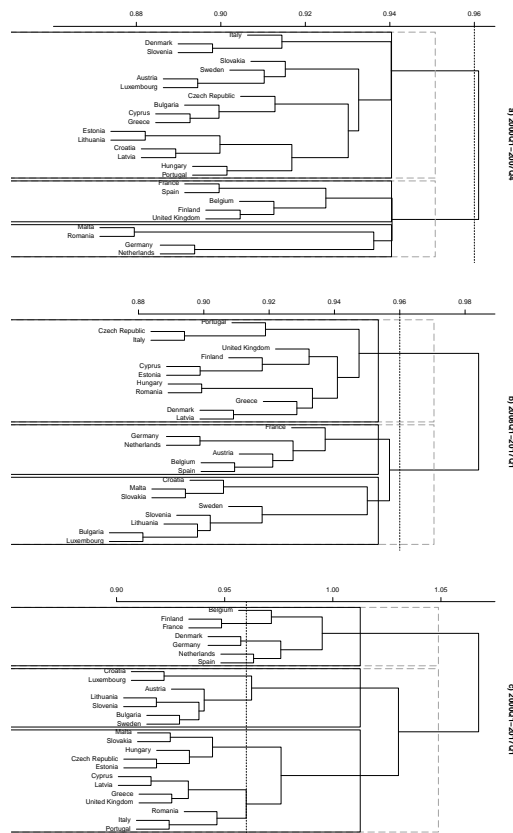


Figure 1: GDP time series clustering. Quarterly data for the EU Member States (Poland and Ireland excluded because of the data availability), millions of euro.

3.2 Clustering Stability

We begin empirical testing with analysis of the time stability of the CDM measure. To this end, we compare time series of the GDPs of the EU Member States, covering three periods: 2000Q1–2007Q4, 2008Q1–2017Q1, and the complete period of 2000Q1–2017Q1 (Figure 1).

The first conclusion is rather obvious: the longer the time series are, the less similarity emerges. This phenomenon is indicated by the dotted line.

A related remark is that the financial crisis reversed processes of economic integration within the EU (both periods are of similar length). This is consistent with the findings of Ahlborn and Wortmann (2018); Belke, Domnick, and Gros (2017); Gächter, Riedl, and Ritzberger-Grünwald (2012), who employed different synchronization measures (respectively: correlation, panel regressions, and nonparametric regressions; correlation; and fuzzy clustering).

The employed similarity method allows us to distinguish two main groups of countries (grey outer rectangles in Figure 1c): the “core”, consisting of France, Germany, the Netherlands, Austria, Belgium, and Spain; and the “periphery”. This structure was slightly different before and after the financial crisis (respectively, Figures 1a and 1b).

Hierarchical representation of the distances enables changing the level of clustering

(black inner rectangles). On this lower level, in the first period (2000–2007) the countries formed four quite similar groups consisting of: (a) Italy, Denmark, and Slovenia; (b) Germany, the Netherlands, Malta, and Romania; (c) France, Spain, UK, Finland, and Belgium; (d) the remaining 14 Member States.

In the second period, generally speaking, after the financial crisis, countries are less similar and it is more appropriate to cluster them into three groups: (a) the “core”, consisting of France, Germany, the Netherlands, Austria, Belgium, and Spain; (b) the “semiperiphery”, consisting of Croatia, Malta, Slovakia, Sweden, Slovenia, Lithuania, Bulgaria, and Luxembourg; (c) the “periphery”, consisting of the remaining 12 countries, including Greece, Italy, Portugal, and the United Kingdom. This structure holds also for the full time series.

These results are similar to (countries in common typeset in boldface):

- Belke et al. (2017), who distinguished **Finland, France, Germany**, Austria, and **the Netherlands** as the core countries;
- Papageorgiou, Michaelides, and Milios (2010), where the core countries group in 2000–2009 consists of Sweden, Portugal, **Germany, France, Spain, Belgium**, Denmark, Austria, and **the Netherlands**;
- Ahlborn and Wortmann (2018), who grouped Austria, **Belgium, Denmark, Finland, France, Germany**, Hungary, Ireland, **the Netherlands**, Norway, Poland, Portugal, **Spain**, Sweden, Switzerland, and the United Kingdom as the core countries in 1996Q1–2015Q4.

To validate the usefulness of the proposed method in the analysis of the structure of an economy, we have selected variables used to calculate the gross domestic product (GDP) and GDP itself.

3.3 Analysis of National Economies

According to Eurostat (2013, p. 273), there are three approaches to calculating GDP:

production approach GDP is the sum of the gross value added of the various institutional sectors or the various industries plus taxes and minus subsidies on products (which are not allocated to sectors and industries);

expenditure approach GDP is the sum of final uses of goods and services by resident institutional units (final consumption of general government, households, and NPISH and gross capital formation) plus exports and minus imports of goods and services;

income approach GDP is the sum of uses in the total economy generation of income account (compensation of employees, taxes on production and imports minus subsidies, gross operating surplus and mixed income of the total economy).

Taking into consideration economic theory and how GDP is calculated, the following distances between variables should be small in all of the analysed countries:

- *GDP and gross value added*,
- *compensation of employees and household and NPISH final consumption expenditure*.

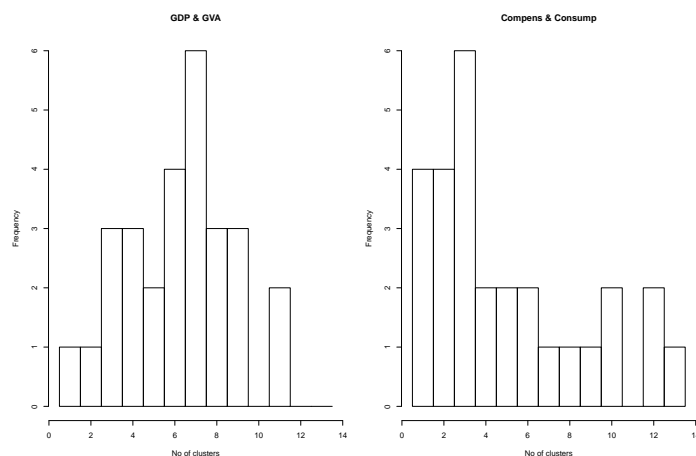


Figure 2: Histograms of clustering stages at which (left) *GDP* and *gross value added*, or (right) *compensation of employees* and *household and NPISH final consumption expenditures* become clustered together for each of the 28 EU Member States.

To verify these relations, we performed an agglomerative clustering on a set of 14 macroeconomic variables (for a full list see Figure 3) that include these four, for each of the 28 EU Member States in the period of 2002Q1–2016Q4. The computation begins with all the variables in separate clusters, and gradually links them together until the two variables of the pair are in one cluster. The number of clusters at this stage is indicative of how similar are the variables of the pair in comparison with the others: the earlier they are clustered together, or the higher the final number of clusters, the more similar they are.

As a rule of thumb, we propose that if compared variables belong to the same group when three or more clusters are created, then the method proposed in this paper is positively verified (confront Figures 2 and 3).

In the case of *GDP* and *gross value added*, there were only two countries for which these variables were linked only in the two last steps of cluster merging. Note that such a situation should raise questions about data quality.

In the case of *compensation of employees* and *household and NPISH final consumption expenditures*, the relation is not so straightforward, and therefore the relative distances are larger. Nevertheless, in half of the analysed countries these variables were connected when the number of clusters was more than three.

In our opinion, these results confirm that the proposed method is consistent with SNA methodology and economic theory.

Finally, we selected four countries (Germany, France, Italy, and Spain) for more detailed verification. Relations between variables are compared in Figure 3. As expected, every country has its unique structure.

Nevertheless, the macroeconomic indicators in Spain, Germany, and Italy form three main clusters, while in the case of France it is more natural to distinguish four main groups.

A more detailed look into the selected countries reveals that all three approaches to the calculation of GDP are generally preserved. This also confirms the robustness of the proposed method.

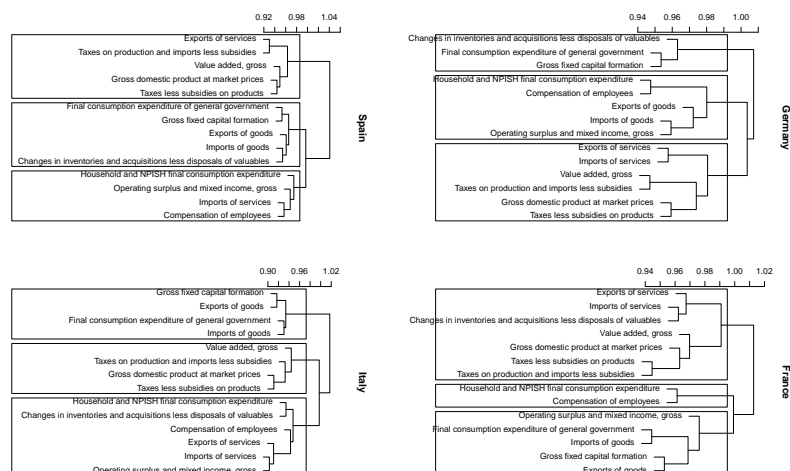


Figure 3: Structure of GDP and its components in Germany, France, Italy, and Spain. 2000Q1–2017Q1 period, millions of euro.

4 Concluding Remarks

The aim of this paper was to verify usefulness of the time series clustering method for macroeconomics and economics policy research, and to develop the most suitable methodology.

After the evaluation of 24 dissimilarity measures, the CDM measure was assessed as the most suitable for our purposes. It most effectively prefers time series with similar frequencies over similar magnitudes of volatility. We assume that likeliness of frequencies could suggest causality.

This assumption and the usefulness of the proposed method is confirmed in the analysis of the distances between aggregated time series of the EU Member States. Results confirm that CDM preserves relations from the national accounts equations as well as from economic theory. It is also robust for manipulations in the length of time series.

It could be therefore recommended for a variety of macroeconomic research topics.

References

- Ahlborn, M., & Wortmann, M. (2018, January). The core–periphery pattern of european business cycles: A fuzzy clustering approach. *Journal of Macroeconomics*, 55, 12–27. Retrieved from <http://hdl.handle.net/10419/152248> doi: 10.1016/j.jmacro.2017.08.002
- Belke, A., Domnick, C., & Gros, D. (2017). *Business Cycle Synchronization in the EMU: Core vs. Periphery* (Working Paper No. 38). GLO Discussion Paper. Retrieved from <http://hdl.handle.net/10419/156158> doi: 10.4419/86788765
- Croux, C., Forni, M., & Reichlin, L. (2001). A Measure Of Comovement For Economic Variables: Theory And Empirics. *The Review of Economics and Statistics*, 83(2), 232–241. Retrieved from <https://ideas.repec.org/a/tpr/restat/v83y2001i2p232-241.html> doi: 10.1162/00346530151143770

- Eurostat. (2013). *European system of accounts – ESA 2010*. Luxembourg: Author. Retrieved from <http://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-02-13-269> doi: 10.2785/16644
- Focardi, S. M., & Fabozzi, F. J. (2004). Clustering economic and financial time series: Exploring the existence of stable correlation conditions. *Finance Letters*, 2(3), 1–9. Retrieved from <https://www.researchgate.net/publication/2547390>
- Gächter, M., Riedl, A., & Ritzberger-Grünwald, D. (2012). Business Cycle Synchronization in the Euro Area and the Impact of the Financial Crisis. *Monetary Policy & the Economy*, 12(2), 33–60. Retrieved from <https://ideas.repec.org/a/onb/oenbmp/y2012i2b2.html>
- Haan, J., Inklaar, R., & Jong-A-Pin, R. (2008). Will business cycles in the Euro Area converge: a critical survey of empirical research. *Journal of Economic Surveys*, 22(2), 234–273. Retrieved from <http://hdl.handle.net/11370/a5248e69-19bb-4057-9284-0876eac01ac2> doi: 10.1111/j.1467-6419.2007.00529.x
- Hamilton, J. D. (2017, May). *Why You Should Never Use the Hodrick-Prescott Filter* (Working Paper No. 23429). National Bureau of Economic Research. Retrieved 2017-07-20, from <http://www.nber.org/papers/w23429> doi: 10.3386/w23429
- Keogh, E., Lonardi, S., & Ratanamahatana, C. A. (2004). Towards Parameter-free Data Mining. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 206–215). New York, NY, USA: ACM. Retrieved 2017-08-17, from <http://doi.acm.org/10.1145/1014052.1014077> doi: 10.1145/1014052.1014077
- Keogh, E., Lonardi, S., Ratanamahatana, C. A., Wei, L., Lee, S.-H., & Handley, J. (2007, February). Compression-based data mining of sequential data. *Data Mining and Knowledge Discovery*, 14(1), 99–129. Retrieved from <https://link.springer.com/article/10.1007/s10618-006-0049-3> doi: 10.1007/s10618-006-0049-3
- Papageorgiou, T., Michaelides, P. G., & Milios, J. G. (2010, September). Business cycles synchronization and clustering in Europe (1960–2009). *Journal of Economics and Business*, 62(5), 419–470. Retrieved 2017-07-20, from <http://www.sciencedirect.com/science/article/pii/S0148619510000469> doi: 10.1016/j.jeconbus.2010.05.004

Value type, perturbation	Global		Number of times ranked			
	maximum		first		top five	
Absolute None	ncd	0.66	ncd	68	ncd	81
	<u>cdm</u>	0.85	dtw	12	dtw	72
	dtw	1.44	pdca	1	<u>cdm</u>	62
			(<u>cdm</u>	0)	pdca	50
Absolute Scaling	<u>cdm</u>	1.33	pdca	50	<u>cdm</u>	81
	ncd	2.07	<u>cdm</u>	22	ncd	74
	cor	2.20	int.per	8	pdca	69
			acf	1	int.per	56
				cor	49	
Absolute Shifting	<u>cdm</u>	1.18	pdca	36	<u>cdm</u>	64
	ncd	1.60	<u>cdm</u>	19	= pdca	64
	cor	2.20	spec.llr	13	ncd	50
			4 more	≤ 5 ea.	int.per	49
					per	45
Absolute Noise	<u>cdm</u>	1.26	dtw	54	dtw	67
	cid	1.60	<u>cdm</u>	13	cid	62
	dtw	1.63	cid	5	<u>cdm</u>	50
	4 more	< 2	5 more	≤ 4 ea.	infnorm	49
					wav	43
Absolute All	<u>cdm</u>	1.33	<u>cdm</u>	49	<u>cdm</u>	81
	ncd	2.09	int.per	26	ncd	79
	cor	2.20	acf	4	int.per	59
			pdca, pacf	1 ea.	cor	57
					ccor	54
					pacf	44
Percentage None	ncd	0.85	ncd	63	ncd	81
	<u>cdm</u>	0.95	dtw	13	dtw	73
	dtw	1.19	pdca	5	<u>cdm</u>	67
			(<u>cdm</u>	0)		
Percentage Noise	<u>cdm</u>	1.11	dtw	33	<u>cdm</u>	78
	ncd	1.41	<u>cdm</u>	20	dtw	69
	dtw	1.50	per, pacf	11 ea.	ncd	55
	6 more	< 2	3 more	2 ea.	cid	41

Table 1: Summary of experiment described in Section 2. Refer to the main text for details. In the last column, only listed are measures that rank in the top five spots more than 40 times, i.e. for more than half of the possible combinations of warp factor and delay.