IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 7 December 2022, accepted 26 January 2023, date of publication 2 February 2023, date of current version 8 February 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3241850

METHODS

Detecting Regimes of Economic Growth With Fuzzy Concept-Based Models

JAKUB BARTAK [1] AND AGNIESZKA JASTRZĘBSKA [2]
[1]Department of Process Management, Cracow University of Economics, 31-510 Kraków, Poland
[2]Faculty of Mathematics and Information Science, Warsaw University of Technology, 00-662 Warsaw, Poland
Corresponding author: Jakub Bartak (bartakj@uek.krakow.pl)

This work was supported by the National Science Centre under Grant 2019/35/D/HS4/01594 (Decision No. DEC-2019/35/D/HS4/01594).

ABSTRACT The paper presents a new method to detect economic growth regimes based on time-series similarity analysis. The procedure builds upon a fuzzy concept-based model, the Jaccard Index, and a hierarchical clustering procedure. Having introduced the proposed method, we apply it to the time series of economic growth and its proximate factors. Further, we study the stability of the obtained results under different parameters. The proposed procedure allows us to detect groups of countries that share a common growth performance and akin growth mechanisms. Simultaneously, the method let us identify countries that escape any fixed classifications.

INDEX TERMS Concept-based models, economic growth, growth regime, fuzzy sets.

I. INTRODUCTION

Understanding the vast differences in economic growth performance and its mechanisms and underlying reasons is one of the fundamental topics in economics. An important strand of research within growth economics deals with the existence of growth regimes, which help explain observed heterogeneity in the growth processes.

The problem of economic growth regimes has an important meaning for macroeconomic theory and empirical studies. The existence of different growth regimes contradicts many theoretical predictions that model growth process as a stable path, accompanied by short-term fluctuations. Moreover, it gives support to models that allow for multiple steady-states, where economies are attracted to different basins of convergence [1], given their initial incomes, the deep determinants of growth [2], or the transitions to prosperity through various growth stages [3]. Simultaneously, the research on economic growth regimes may also have substantial policy implications. As pointed out in [4], detecting distinct growth processes that result in diverging convergence clubs necessitates extensive interventions to address catching-up barriers. Moreover, the identification of growth regimes can aid in the formulation of growth enhancing policies that are regime-specific, and thus - better suited for individual countries. With this study, we hope to contribute to the literature by proposing a novel procedure for identifying growth regimes that is more comprehensive than existing methods.

There exist two distinct approaches to the research on growth regimes. In the first one, the growth regime is understood as a set of circumstances, that affect the mechanisms of the GDP (Gross Domestic Product) growth process [5]. In such a view, the regime is associated with a set of deep and long-lasting determinants of development possibilities [2]. As a result, each regime is characterized by a different relationship between proximate growth determinants and the growth itself. In such an approach, the methods to detect growth regimes rely on identifying the differences in these relationships. Thus, a regime is detected on the basis of the analysis of the growth time-series generation process. In the second approach, a growth regime is understood as a temporal characteristic of the growth time series. The regime is defined as a model or situation of economic performance that qualitatively differs from the others [6]. Within this perspective, the growth regime is strictly associated with the GDP time series rather than with its determinants. Within this approach, the growth regime is hardly distinguishable from a growth pattern, as defined in [7]. Nevertheless, the silent assumption that accompanies the search for growth regimes within this strand of research is that a different set

The associate editor coordinating the review of this manuscript and approving it for publication was Kostas Kolomvatsos.

of structural determinants generates qualitatively different growth
performance.

Thus, the main difference between the two approaches is the starting point of the analysis. In the first one, the starting point is the relationship between growth and its determinants, while the second is the growth performance itself. Simultaneously, the general notion of the 'regime' identified in both approaches is the structural difference in the growth mechanism.

In the current paper, we propose to identify growth regimes with a novel, multi-dimensional concept-based model. The proposed method allows operating in the two realms of regime definitions. It determines growth regimes based on growth performance and different growth mechanisms, understood as the temporal relationship between growth and its proximate factors. In other words, the model helps to identify countries (or growth episodes) that are similar in terms of performance and in terms of the proximate driving forces of such performance. In this sense, the growth regime is defined not only as a growth pattern but as a pattern explained by a given set of conditions. Notably, the pattern itself and the growth conditions can vary over time and be non-linear.

On the methodological level, the new approach brings forth ideas from the fuzzy sets theory, which is used to build an initial data model. Subsequently, classic clustering and regression analysis algorithms are involved in extracting patterns from the fuzzy concept-based model. The proposed method is a fusion between soft and crisp computing technologies. The proposed approach is a continuation of our research on imprecise knowledge representation schemes applied to model temporal data addressed in [8] and [9].

The rest of the paper is structured as follows. Section II reviews existing literature on the topic, Section III describes the proposed model, Section IV provides the application of the model and the corresponding results. Further, we study robustness of the main results in Section V. Conclusions are stated in Section VI.

II. LITERATURE REVIEW

The existing methods aimed at discovering growth regimes can be grouped into two strands of literature. In the first one, a growth regime is understood as the overall environment influencing the behavior of economic agents, connected with deep determinants of growth, such as economic institutions, which shape the relationship between growth and its proximate determinants, i.e. accumulation of production factors and economic policy [5]. Searching for GDP growth regimes under this approach depends on statistical procedures aimed at discovering distinct clusters of economies, which are differentiated in terms of the relationship between growth determinants and the growth itself.

In an early attempt to achieve this aim researchers tried to exogenously define sub-populations of countries on the basis of a qualitative assessment of their initial conditions [10]. Considering the arbitrariness of such attempts, researchers moved towards more data-driven approaches. Reference [11] proposed an application of a regression tree to search for such a split of the data, that would maximize the fit of the cross-section growth regression model. This approach has been developed in [12], where a regression tree was applied to identify threshold levels of a set of candidate variables that cluster economies into distinct groups. The outcomes of these studies revealed that the coefficients of growth determinants in the linear models vary across groups. In this sense, the groups could be interpreted as countries that follow distinct growth regimes. A further extension was proposed by Fiaschi et al. [2], who compared Generalized Additive Models of economic growth, and identified a sampling split, that is the most informative (as measured by Akaike Information Criteria). Such a procedure allowed to account for the fact that the relationship between growth and its determinants may be non-linear.

Another method applied to the task of regime detection is a finite mixture model. The finite mixture model assumes that the observed conditional distribution of the variable of interest comes from several unobserved sub-populations. These latent sub-populations are discovered by estimating the probabilities of falling into a given class. In its application to economic growth regime problems, the modeled distribution of economic growth rates is conditional to the class membership and the set of independent variables – typically to the standard set of growth factors. An essential feature of the model is that the classes and the parameters of the growth regression, are estimated jointly within the model and may vary across classes (regimes). Thus, the latent classes may reflect the differences in the marginal effects of the growth factors. The finite mixture model has been applied, among others, by Owen et al. [5], Di Vaio and Enflo [13], Flachaire et al. [14], and Liu et al. [15].

In the second strand of literature, a growth regime is understood as a temporal characteristic of the growth time series. Within this approach, the effort of researchers is concentrated on pinpointing the episodes of high/low growth. As proposed by Brida et al. [6], one possible solution is to use a symbolic representation of the growth rates, which facilitates classifications based on external thresholds. A growth rate exceeding the given threshold is classified as ''high'', and ''low'' in another case. A possible extension to this technique is growth rate quantization. Other approaches use Markov-switching models, where the growth process is interpreted as a series of transitions between growth states. A growth state is modeled as an auto-regressive process, which is described with state-specific coefficients that reflect the growth rate and its volatility typical to a given state. Then, a growth state is interpreted as a growth regime. In this way, the growth regimes were studied by Jerzmanowski [16] where the author focused on the impact of economic institutions on the probabilities of transition between regimes. In another paper, Kerekes [17] clustered economies based on similarities in transition probabilities and described clusters with a series of economic variables. Finally, Morier and Teles [18] extended the previous work with the application of the time-varying Markov-switching model and analyzed the regime-specific relationship between investment and growth.

To sum up, we can state two generalizations:

- The current state of the art enables to group countries based on distinct growth performance. This may serve as a starting point for the analysis of the group-specific determinants of GDP growth, as well as the analysis of the factors impacting growth accelerations [19], and growth spells [20], [21]. Within this approach, the grouping itself does not relate to the similarity in the growth processes and mechanisms but rather to the growth performance.
- Another set of methods enables the identification of groups of countries with similar growth processes, i.e., with a similar relationship between growth and its proximate determinants. These methods do not take into account the episodic character of growth [7] and the possibility that countries may experience a change in the structural characteristics (which may impact the growth process itself). Also, the methods based on finite mixture models lack treatment of non-linearities between growth and its factors.

The model we propose below is aimed at merging the existing approaches, i.e., our method can detect growth regimes resulting from the growth performance and the differences in the growth mechanisms.

III. FUZZY CONCEPT-BASED MODEL

A. PRELIMINARY NOTES

The model utilized in this study is a specific variant of a concept-based model delivered in a unified formalism by Jastrzebska et al. [8]. In this study, we have performed the required alterations of this generalized formalism to represent and process data concerning growth regimes. Thus, the novelty of the contribution presented in this paper relies on a creative adaptation of an existing modeling framework.

In our previous work, we have already delivered one particular adaptation of the original concept-based model to economics. The mentioned study by Bartak and Jastrzebska [9] targeted the discussion on evaluating the similarity of transitional growth. The model presented in the current paper differs from [9] in several ways:

- It applies the proposed procedure to a multi-dimensional problem, whereas in [9], the model was applied to two dimensions.
- The concepts are generated with fuzzy c-means algorithm run on multidimensional data instead of the Cartesian product of the c-means algorithm run separately for each variable.
- In the preprocessing stage, the weights attached to the variables impact the concept generation procedure. Specifically, the input variables are normalized to have a mean equal to 0 and a standard deviation equal to 1. Then, the so-normalized data are multiplied accordingly to reflect the assumed weights. In this way, depending on the research objective, the model returns an outcome better suited to the specific research question.
- The model is run not only for countries but also for country-episodes, defined as an episode of economic growth which is not interrupted by any structural break. In this way, we can account for the possibility that a country switches growth regimes along its development path.

B. PROPOSED MODEL ADAPTATION

The procedure can be roughly split into three steps:

1) Concept extraction and membership matrix computation.
2) Similarity evaluation.
3) Exploration of countries' similarities

1) STEP 1: CONCEPT EXTRACTION AND MEMBERSHIP COMPUTATION

On the input to the model, we receive multivariate time series concerning several countries. Eq. (1) specifies the input data format.

$$\begin{array}{l}
\text{country 1:} \begin{cases} \text{variable 1:} & z_1^{11}, z_2^{11}, \ldots, z_{N^1}^{11} \\ \text{variable 2:} & z_1^{12}, z_2^{12}, \ldots, z_{N^1}^{12} \\ \vdots & \\ \text{variable P:} & z_1^{1P} z_2^{1P}, \ldots, z_{N^1}^{1P} \end{cases} \\
\text{country 2:} \begin{cases} \text{variable 1:} & z_1^{21}, z_2^{21}, \ldots, z_{N^2}^{21} \\ \text{variable 2:} & z_1^{22}, z_2^{22}, \ldots, z_{N^2}^{22} \\ \vdots & \\ \text{variable P:} & z_1^{2P} z_2^{2P}, \ldots, z_{N^2}^{2P} \end{cases} \\
\qquad\qquad\vdots \\
\text{country R:} \begin{cases} \text{variable 1:} & z_1^{R1}, z_2^{R1}, \ldots, z_{N^R}^{R1} \\ \text{variable 2:} & z_1^{R2}, z_2^{R2}, \ldots, z_{N^R}^{R2} \\ \vdots & \\ \text{variable P:} & z_1^{RP} z_2^{RP}, \ldots, z_{N^R}^{RP} \end{cases}
\end{array} \tag{1}$$

Let us use P and R to denote the number of variables and countries, respectively. N^j, $j = 1, \ldots, R$ is the length of each time series for country j, and it is the same for all time series for one country. The lengths may differ between various countries. $P, R, N^j \in \mathbb{N}_+$. $z_i^{jk} \in \mathbb{R}$ is an ith element of a series assigned to country j, variable k, $i = 1, \ldots, N^j, j = 1, \ldots, R, k = 1, \ldots, P$. Let us also assume that variables are sorted. That is, the kth variable is the same for all countries in the dataset. Let us note that time series are collected at equal time intervals.

The first step of the procedure relies on concept extraction. This step results in the creation of a concept-based model of the input data. Concepts can be implemented using various formalisms. In this study, like in previous ones, we use fuzzy sets. The role of the concepts is to aggregate and represent the raw data. Thus, fuzzy sets are a good choice. One can conveniently extract fuzzy sets using the fuzzy c-means algorithm proposed by Bezdek [22].

Concepts are extracted for input data organized in a manner that each column corresponds to one variable and each row corresponds to one observation. The fuzzy c-means algorithm does not consider temporal dependencies in the data. That is, the order of rows is not important. In our case, input data is formed by concatenating the data concerning all the countries. It is performed so that corresponding variables are concatenated with each other. For example, the first variable after concatenation is given as follows: $z_1^{11}, z_2^{11}, \ldots, z_{N^1}^{11}, z_1^{21}, z_2^{21}, \ldots, z_{N^2}^{21}, \ldots, z_1^{R1}, z_2^{R1}, \ldots, z_{N^R}^{R1}$. The number of countries is R and the length of time series for a country j is N^j. Thus, after concatenation, we obtain $N^1 + N^2 + \ldots + N^R$-long variables. For the convenience of notation let us denote this sum as N, that is, $N = N^1+N^2+\ldots+N^R$. Suppose we collect concatenated variables into a single matrix necessary for the fuzzy c-means algorithm. In that case, we obtain an $N \times P$ matrix of real values, where each column corresponds to one variable. Let us denote this matrix as $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]$, where $\mathbf{x}_i = [x_{i1}, x_{i2}, \ldots, x_{iP}]$, $x_{ik} \in \mathbb{R}$ is a single row of the matrix $\mathbf{X}$.

Fuzzy c-means is an iterative clustering algorithm. It starts with an initialization of the location of centroids (interpreted as fuzzy sets), which can be performed with something as simple as a random choice. Subsequently, the algorithm adjusts the location of the existing centroids by minimizing an objective function given in Eq. (2).

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} \mu_{ij}^m \|\mathbf{x}_i - \mathbf{v}_j\|^2. \tag{2}$$

In Eq. (2), centroids are denoted as $\mathbf{v}_j, j = 1, \ldots, C$, C is the number of centroids. We must specify C before the algorithm is launched since in consecutive iterations the number of centroids does not change. Please note that $\mathbf{v}_j$ is written in bold font since it is a vector of real values. Centroids are located in the space of variables used to execute the algorithm.

μ_{ij} is computed according to Eq. (3),

$$\mu_{ij} = \frac{1}{\sum\limits_{k=1}^{C}\left(\frac{\|\mathbf{x}_i-\mathbf{v}_j\|}{\|\mathbf{x}_i-\mathbf{v}_k\|}\right)^{2/(m-1)}} \tag{3}$$

and

$$\mathbf{v}_j = \frac{\sum\limits_{i=1}^{N} \mu_{ij}^m \cdot \mathbf{x}_i}{\sum\limits_{i=1}^{N} \mu_{ij}^m}, \tag{4}$$

where m is a fuzzification coefficient $m \in (1, \infty)$, defining the degree of fuzziness of the solution. $m \to 1$ results in the behavior analogous to the behavior of the crisp k-means clustering algorithm. Increasing m leads to more fuzzy assignments to the centroids in a model. This parameter is often set to 2, the value recommended by Hathaway and Bezdek [23].

The algorithm terminates its operation after a predefined number of iterations is reached or when a difference between subsequent locations of the centroids does not change more than a specified threshold ε. In this case, we assume that the centroids have stabilized, and we exit the procedure. The outcome of the fuzzy c-means algorithm is a collection of centroids (fuzzy sets) and a membership matrix $[\mathbf{M}] = \mu_{ij}$, $i = 1, \ldots, N, j = 1, \ldots, C$ specifying the membership of each data point from the input dataset to each fuzzy set. Please note that formulas Eq. (3) and (4) ensure that for each data point, the respective memberships add up to 1. More formally, $\forall_i \sum_{j=1}^{C} \mu_{ij} = 1$.

Fuzzy c-means can be susceptible to outlying observations. Therefore, it is recommended to remove them before the initiation of the procedure. After outlier removal, we standardize the data employing the well-known formula $x' = \frac{x-\bar{x}}{\sigma}$ with standard deviation σ and mean value $\bar{x}$ of a variable.

The next step is concept extraction. As matrix $\mathbf{X}$ from Eq. (2)–(4), we take concatenated and preprocessed multivariate data given originally in Eq. (1). The number of concepts C is a model parameter, which can be set individually. The experiments presented in this paper use $C = 100$. The number of concepts should be large enough to reflect various regularities present in the data, but at the same time, too many concepts may be redundant. Since the model anyway is to be interpreted by an expert, we suggest testing several values of this parameter.

The advantage of a concept-based knowledge representation is that we can interpret the values (locations) of the concepts using expert knowledge and treat concepts as knowledge aggregates. In other words, concepts represent "typical" raw data points. This interpretation is valid as fuzzy c-means is essentially a clustering algorithm that aims at the detection of groups of observations.

2) STEP 2: SIMILARITY EVALUATION

The subsequent step of the procedure aims at the similarity evaluation of data concerning different countries. Thus, for data concerning each country individually, we compute membership values for each one of the concepts extracted in the previous step. The formula for membership evaluation is given in Eq. (3). Subsequently, we sum the belongingness to each concept. In the end, for each country, we obtain a single, C-element vector in the form $\mathbf{a}_i = [a_{1i}, \ldots, a_{Ci}]$, where i is the country index and $a_{ji} = \sum_{k=1}^{N_i} \mu_{kj}$ is the sum of memberships for all data points concerning country i to the concept j. For the sake of simplicity of notation, let us assume that N_i is the number of membership values being summed. This sum is a real number, and thus, after computing $\mathbf{a}_i$ for each country, we normalize these values to the [0, 1] interval by employing the standard min-max normalization (that is $x' = \frac{x-\min(x)}{\max(x)-\min(x)}$).

One may say that given the unequal length of the time series concerning different countries the sum (and normalization) operations may lead to some distortions in the perception of the regularities present in the data. Far from it, we want to utilize knowledge about the shorter series since, typically, the lack of prior data points is also something distinctive for a given country.

Subsequently, we are ready for similarity evaluation, which is performed with the use of the Jaccard index. In its generalized form, it is given as follows:

$$SIM\_Jaccard(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}, \quad (5)$$

where A and B are sets. When applied to normalized real vectors, we can formulate it as follows:

$$SIM\_Jaccard(\mathbf{u}, \mathbf{v}) = \frac{\sum_{i=1}^{N} \min(u_i, v_i)}{\sum_{i=1}^{N} \max(u_i, v_i)} \quad (6)$$

$\mathbf{u}, \mathbf{v}$ are vectors $\mathbf{v} = [v_i], \mathbf{u} = [u_i], i = 1, \ldots, C, u_i, v_i \in [0, 1]$. We obtain $R \times R$ similarity matrix with rows and columns indexed with country names.

Algorithm 1 outlines the procedure in the form of a pseudocode.

Algorithm 1 The Procedure at a Glance

```
Data: Multivariate time series concerning R countries
      formalized in Eq. (1).
Result: M, R × R similarity matrix.
Choose C, the number of centroids;
Concatenate corresponding variables from each
  country;
Remove significant outliers from the concatenated
  data (values beyond mean plus/minus triple standard
  deviation);
Run fuzzy c-means for concatenated multivariate
  dataset to obtain C centroids;
for i = 1, ..., R do
    Initialize C-elements vector, denote it as u_i;
    for c = 1, ..., C do
        Compute membership values for the ith
          country to the cth centroid;
        u_i[c] = sum of membership values
          corresponding to the concept c.;
    end
end
for i = 1, ..., R do
    Normalize u_i to [0, 1] using min-max
      normalization;
end
for i = 1, ..., R do
    for j = 1, ..., R do
        M[i, j] = SIM_Jaccard(u_i, u_j), according to
          the formula Eq. (6);
    end
end
```

3) STEP 3: EXPLORATION OF COUNTRIES' SIMILARITY

Having obtained a similarity matrix, we can perform further data exploration. This exploration aims to the identification of groups of countries deemed to be similar according to the intermediate data representation produced with Steps 1 and 2 outlined above. To automate this task, we propose to apply hierarchical clustering.

Hierarchical clustering generates a tree-based representation of a given input data set or a similarity matrix such that observations within a given branch are deemed to be more similar to each other than observations placed in a different branch. The algorithm builds a tree (typically a binary tree), in whose root we have all observations, and in leaves we have single observations. Splits are defined based on the similarity of observations that end up in created nodes. The outcome of the processing is conveniently visualized with the so-called dendrogram. An expert decides where to perform a cut, that is, how many groups should be identified. The lower the cut, the more clusters we generate. The decision about the number of clusters to be generated in the application, as addressed in this study should be made manually by looking at the created dendrogram.

TABLE 1. Variables' averages by growth regime. The model with no weights.

cluster	n	growth rate	income	human capital	population growth rate	capital formation
1	2221	2.14	8.5	1.94	2.14	0.21
2	1369	2.88	9.48	2.42	1.27	0.25
3	1133	2.48	9.88	2.92	0.72	0.29
4	2304	1.24	7.47	1.45	2.55	0.15

IV. DATA AND RESULTS

The study uses an unbalanced panel with annual observations over the period 1951–2017. From the original database (PWT database version 9.1 by [24]), the countries with less than 30 observations were filtered out, and we deleted the outlying observations. This procedure returned a data set comprising 122 countries with a total number of observations equal to 6929. The following variables were used as input in the model: the rate of per capita GDP growth, the natural logarithm of income, human capital, the share of gross capital formation (at current Purchasing Power Parities), and the rate of population growth. Thus, the set of variables includes the standard variables of a Solowian-type growth model.

At first, we turn to the result obtained for the model with no weights. As evidenced in Table 1, the model returned three clusters with fairly similar average growth rates (Clusters 1 to 3) and one group of countries that experienced stagnation throughout the analyzed period (cluster 4). Such a sample split gives interpretation opportunities since the clustering revealed that similar growth rates were obtained in different conditions. Countries in cluster 1 were able to develop despite the rapid population growth rate. Compared to stagnating countries in cluster 4, this group (although relatively poor) is characterized by much higher human capital and investments in fixed assets. Countries in cluster 3, which form the group of relatively prosperous economies with the highest stock of human capital, maintained relatively high

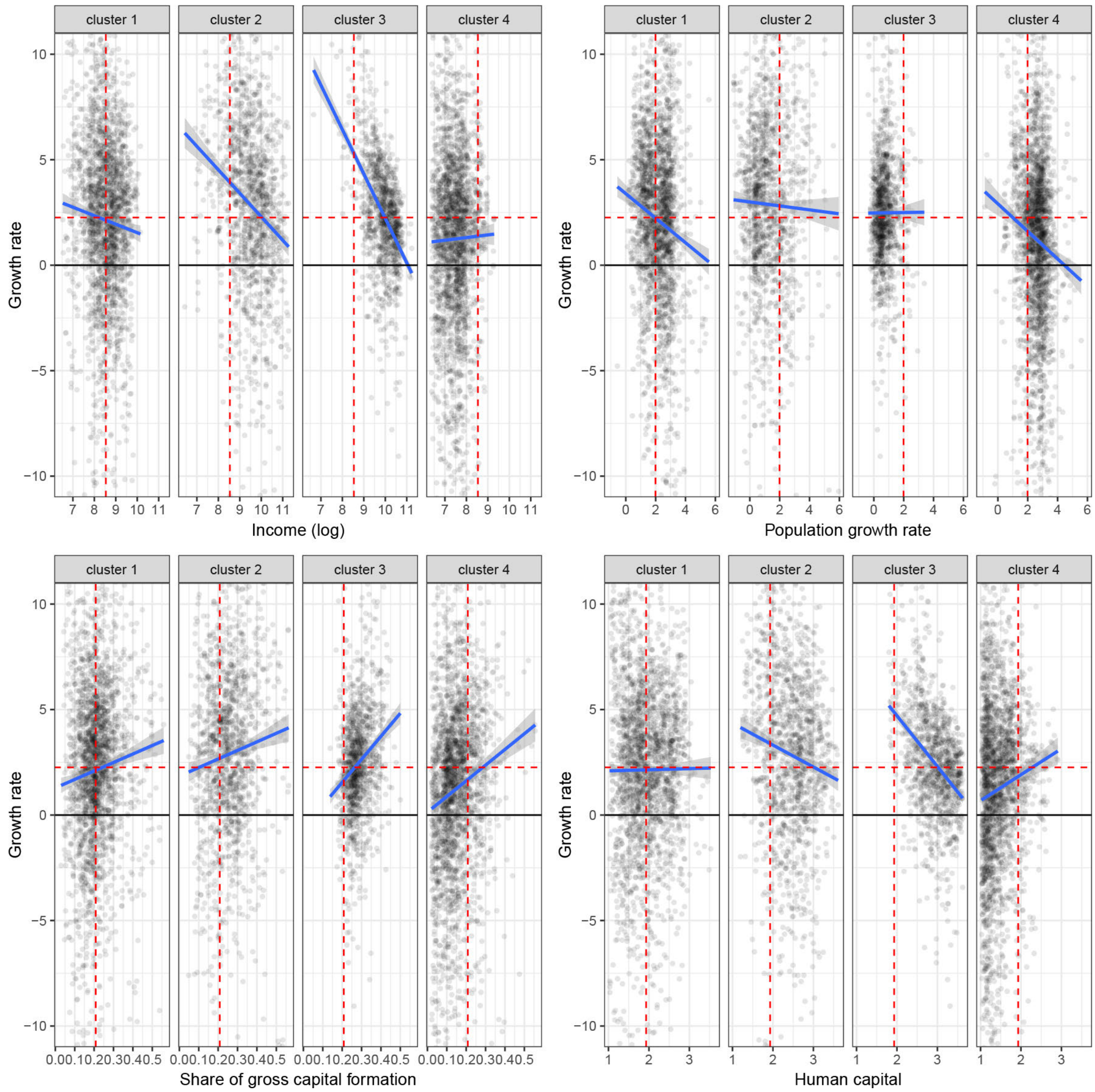

FIGURE 1. Regimes discovered for the model with countries, no weights.

growth. Countries in the second cluster seem to be in between clusters 1 and 3. Regarding the relationship between growth and its proximate determinants, we can also observe some compelling insights (Figure 1). First, the relationship between income and growth is positive in cluster 4. This contrasts the negative relationship (in line with the convergence hypothesis) for countries grouped in clusters 1 to 3. Furthermore, a positive correlation is found between human capital and growth in the case of clusters 1 and 4, whereas the opposite is evidenced in clusters 2 and 3. Finally, the high population growth rate negatively correlates with GDP growth in all clusters except cluster 3, which may point to the aging problems in well-developed countries.

Although the above-presented procedure revealed some insights, it also has disadvantages. Namely, the groupings are based chiefly on the level of crucial variables (which are correlated to some degree) and mimic the common knowledge

TABLE 2. Variables' averages by growth regime. The model with weights (0.6 to growth rate, 0.1 to the remaining variables).

cluster	n	growth rate	income	human capital	population growth rate	capital formation
1 (catching-up)	2505	2.51	8.55	1.96	2.06	0.22
2 (poor)	1480	0.79	7.64	1.46	2.73	0.17
3 (rich)	2101	2.59	9.7	2.72	0.81	0.27
4 (developing)	941	1.58	7.61	1.57	2.46	0.15

about the growth process. Thus, the growth rates within each cluster can be explained by the corresponding levels of its proximate variables. For example, the low rates of growth in cluster 4 can be explained by low capital formation, high population growth, and low human capital. This explanation is rather non-controversial but at the same time, it does not add up to the existing knowledge.

Thus, going forward and trying to discover clusters of countries that vary in economic performance but share some

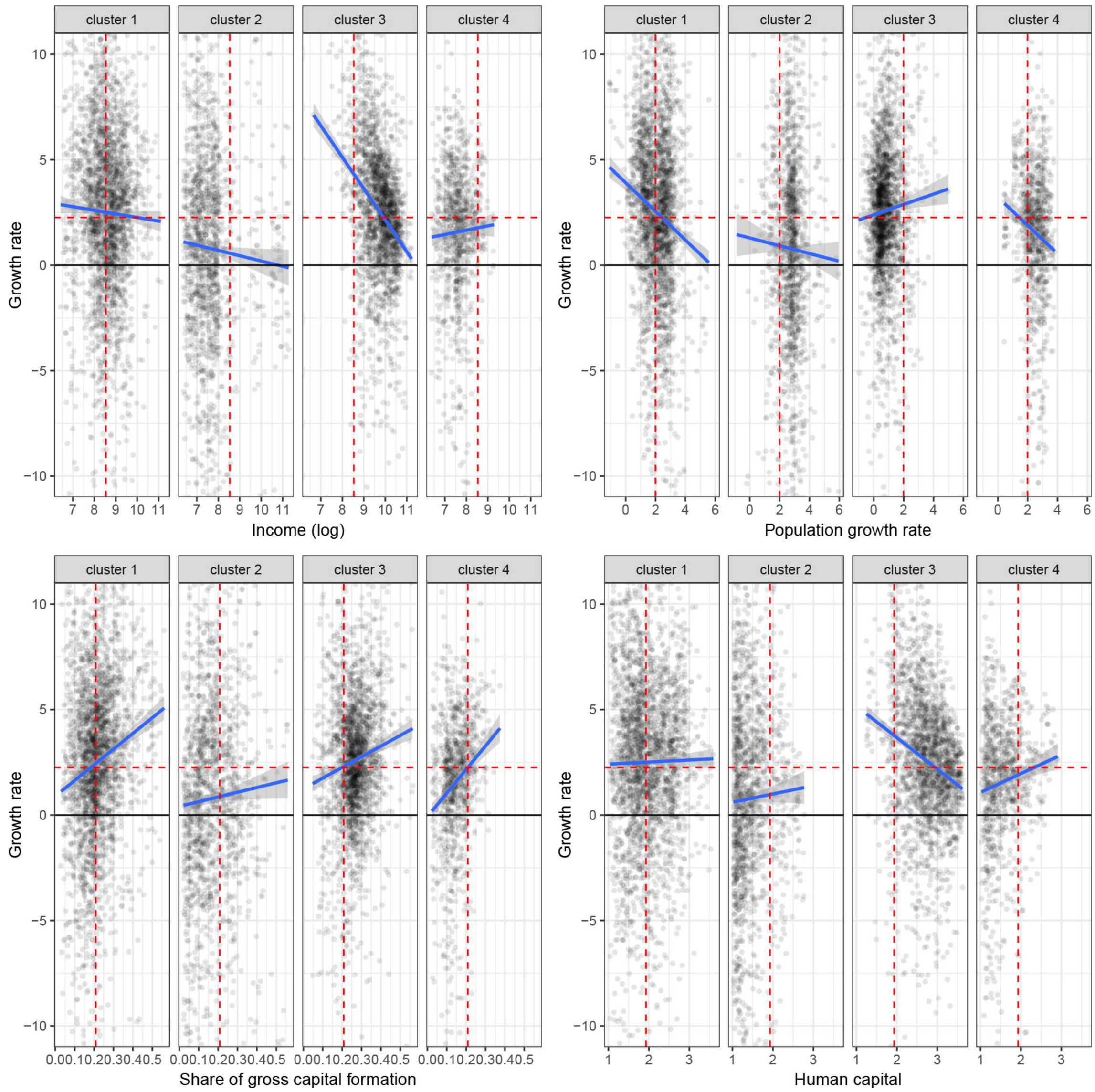

FIGURE 2. Regimes discovered for the model with countries, and with weights (0.6 to growth rate, 0.1 to the remaining variables).

common characteristics of proximate growth determinants, one may use a modification of the proposed procedure and attach weights to the input variables. To study such a procedure, we attached 0.6 weight to the economic growth variable and 0.1 weight to the remaining ones. In this way, we seek clusters that differ mainly in the average growth rates. At the same time, the model seeks patterns and levels of growth factors that constitute a growth regime, but these factors have a lesser impact on the composition of the final clusters.

The outcome of the model with weights is presented in Table 2 and Fig. 2. Given the average growth rates and income level, one can label the clusters as *catching up* (cluster 1), *poor* (cluster 2), *rich* (cluster 3), and *developing* (cluster 4). Striking differences are observed between clusters 2 and 4, i.e., between clusters with similar income, human capital, population growth, and capital formation, but with growth rate in the 4th group doubling the rates in the 2nd group. Such an outcome calls for follow-up studies, which could seek to understand why cluster 4 countries outperformed cluster 2 countries in such a remarkable way. Identification of clusters 2 and 4 can also confirm the need to look for patterns of economic growth in relation to other dimensions of development. The bi-variate relationships presented in Fig. 2 point to correlations similar to the ones discussed in the case of the model with no weights.

Last but not least, while detecting growth regimes, countries may switch regimes due to undergoing structural changes. The basic idea of a regime switch refers to significant economic changes which impact the mechanisms and the dynamics of development. Thus, we avoid any analysis that confuses structural change with temporary shock in growth performance. For this purpose, we detect structural breaks in

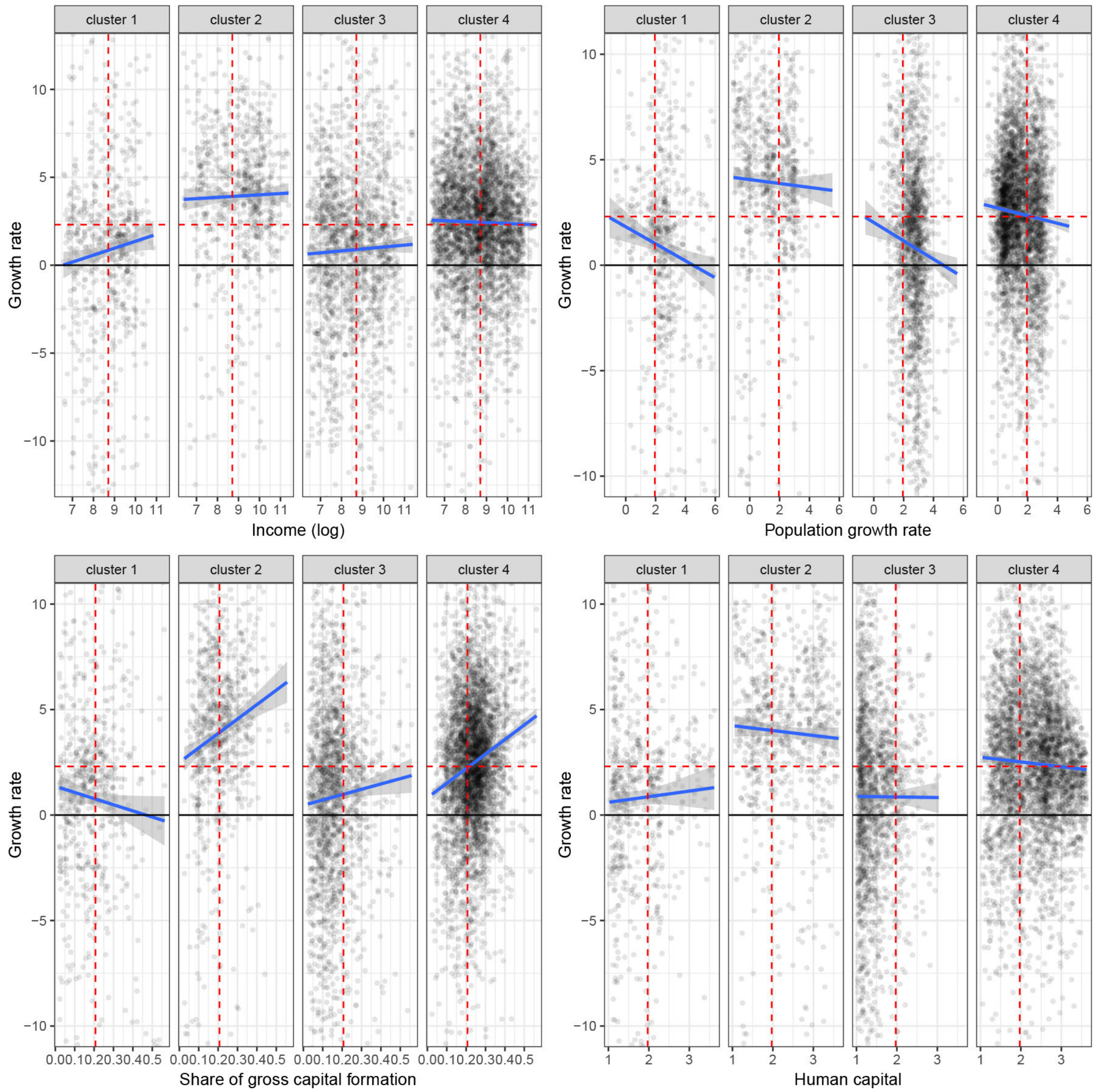

FIGURE 3. Regimes discovered for the model with episodes, and with weights (0.6 to growth rate, 0.1 to the remaining variables).

growth time series with the genetic algorithm. On this basis, we identify episodes with different growth rates and different volatility. We choose the genetic algorithm rather than the Bai-Perron algorithm (reckoned as a standard procedure in this area of study), because it outperforms the Bai-Perron method in two ways [25]:

- It separates episodes of volatile growth from episodes of stable growth, even if both of them have similar average growth rates.
- It is unlikely to detect break-points in case of temporary, short-lasting shocks.

Episodes detected with the genetic algorithm served as objects in the concept-based model, i.e., the average memberships for each concept were calculated for each country-episode. Then, country-episodes were used for computing the similarity matrix and the corresponding clustering procedure.

TABLE 3. Variables' averages by growth regime. Model for growth episodes with weights (0.6 to growth rate, 0.1 to the remaining variables).

cluster	n	growth rate	income	human capital	population growth rate	capital formation	episode length
1	665	0.81	8.61	1.79	2.5	0.19	10.73
2	798	3.93	9.04	2.3	1.34	0.21	14.51
3	1788	0.87	8.53	1.48	2.65	0.16	31.93
4	4044	2.44	8.66	2.33	1.47	0.24	48.14

The clustering results are presented in Figure 3 and Table 3. The process revealed two clusters with periods of stagnation, one cluster of episodes of very high growth, and one cluster of stable and long-lasting growth. Even though the duration of episodes was not used as an input of the model, it turned out to be a vital feature in partitioning our sample. Specifically, the duration of episodes impacts cluster composition through membership-values summing and normalization procedure. As an aftermath, the country-episodes of shorter length are evaluated as different from long country-episodes. This feature is advantageous since the duration of the episodes has important economic meaning [21]. Thus, separating, e.g., clusters 1 and 3 is informative and can serve as a building block for the qualitative or quantitative analysis of the determinants of growth slowdowns and their duration.

V. ROBUSTNESS ANALYSIS

In the current section, we test the stability of the clusters under different values of the model's key parameters. Specifically, we formulate two research questions:

> RQ1: Does the random initialization of the clustering partition matrix affect the model's outcome?
> RQ2: Does the parameter C (number of concepts) impacts the model's output?

To answer RQ1, we run our model 100 times, randomly varying the initial partition matrix each time. To isolate the effects of such randomization, we keep the other parameters constant over 100 runs, i.e. we use $C = 100$, weight assigned to economic growth variable $= 0.6$, and weights assigned to the remaining variables $= 0.1$. To investigate how frequently the models' outcomes change, we report the percentage of classifications into one of our four clusters for each country in our sample (part A of Table 4).

To answer RQ2, we run the model 100 times, increasing the parameter C (the number of concepts) by one each time. We begin with C = 51 and end with C = 150. Similarly to the first experiment, we keep the other parameters constant over 100 runs, i.e. use a fixed initial partition matrix and a weight of 0.6 for economic growth variable and a weight of 0.1 for the remaining variables. We report the percentage of classifications into one of our four clusters in part B of Table 4.

As evidenced in part A of Table 4, in many cases, the countries switch cluster memberships just because of the random initialization of the partition matrix. We can draw a similar conclusion from the experiments with a different number of concepts. On the one hand, such an outcome can be perceived as a marker of disadvantages of the proposed method since the discovered uncertainty cast doubts on the reliability of the groupings. On the other hand, one can also get an interpretation that is more favorable to the model itself. A closer look at those unstable results indicates that countries that switch memberships due to random parameters also tend to change classes due to different number of concepts. Another finding is that the least stable clusters are clusters 1 and 4, which are the closest in terms of the average development levels achieved by their members. Both of these facts can be interpreted as an indication that these ''problematic'' countries are indeed escaping a fixed classification and could be treated as the ones with distinct or changing economic growth regimes. One way of moving forward could be excluding these economies from the subsequent analysis.

TABLE 4. Classification stability.

	Part A Initial partitions				Part B Varying number of concepts			
Cluster	1	2	3	4	1	2	3	4
Albania	99	0	0	1	100	0	0	0
Algeria	53	0	0	47	27	2	0	71
Angola	0	84	0	16	0	82	0	18
Argentina	53	0	0	47	28	2	0	70
Australia	0	0	100	0	0	0	100	0
Austria	0	0	100	0	0	0	100	0
Bahrain	17	58	0	25	7	69	0	24
Bangladesh	2	33	0	65	1	44	0	55
Barbados	1	0	99	0	2	0	95	3
Belgium	0	0	100	0	0	0	100	0
Belize	53	0	0	47	27	2	0	71
Benin	0	43	0	57	0	51	0	49
Bolivia	3	18	0	79	2	18	0	80
Botswana	99	0	0	1	100	0	0	0
Brazil	51	4	0	45	24	5	0	71
Brunei Darussalam	17	58	0	25	7	69	0	24
Burkina Faso	0	84	0	16	0	82	0	18
Burundi	0	100	0	0	0	100	0	0
Cambodia	0	84	0	16	0	82	0	18
Cameroon	0	43	0	57	0	51	0	49
Canada	0	0	100	0	0	0	100	0
Central African Rep.	0	100	0	0	0	100	0	0
Chile	3	0	97	0	2	0	95	3
China	100	0	0	0	100	0	0	0
China, Hong Kong SAR	23	0	77	0	74	0	26	0
China, Macao SAR	99	0	0	1	100	0	0	0
Colombia	51	4	0	45	24	5	0	71
Congo	0	100	0	0	0	100	0	0
Congo, Democratic Rep.	0	100	0	0	0	100	0	0
Costa Rica	51	4	0	45	24	5	0	71
Cote d'Ivoire	0	100	0	0	0	100	0	0
Cyprus	0	0	100	0	0	0	100	0
Denmark	0	0	100	0	0	0	100	0
Dominican Rep.	83	1	0	16	68	0	0	32
Ecuador	51	4	0	45	24	5	0	71
Egypt	3	18	0	79	2	18	0	80
El Salvador	3	18	0	79	2	18	0	80
Eswatini	50	2	0	48	24	2	0	74
Ethiopia	0	84	0	16	0	82	0	18
Fiji	53	0	0	47	27	2	0	71
Finland	0	0	100	0	0	0	100	0
France	0	0	100	0	0	0	100	0
Gabon	53	0	0	47	27	2	0	71
Gambia	0	100	0	0	0	100	0	0
Germany	0	0	100	0	0	0	100	0
Ghana	48	7	0	45	18	10	0	72
Greece	0	0	100	0	0	0	100	0
Guatemala	0	43	0	57	0	51	0	49
Haiti	0	100	0	0	0	100	0	0
Honduras	0	43	0	57	0	51	0	49
Hungary	0	0	100	0	0	0	100	0
Iceland	0	0	100	0	0	0	100	0
India	15	17	0	68	9	35	0	56
Indonesia	99	0	0	1	100	0	0	0
Iran	53	0	0	47	27	2	0	71
Iraq	0	84	0	16	0	82	0	18
Ireland	0	0	100	0	0	0	100	0
Israel	1	0	99	0	2	0	95	3
Italy	0	0	100	0	0	0	100	0
Jamaica	53	0	0	47	27	2	0	71
Japan	0	0	100	0	0	0	100	0
Jordan	53	0	0	47	27	2	0	71
Kenya	0	43	0	57	0	51	0	49
Lao People's DR	100	0	0	0	100	0	0	0
Lesotho	8	88	0	4	1	94	0	5
Liberia	0	84	0	16	0	82	0	18
Luxembourg	0	0	100	0	0	0	100	0
Madagascar	0	100	0	0	0	100	0	0
Malawi	4	93	0	3	1	98	0	1
Malaysia	99	0	0	1	100	0	0	0
Maldives	100	0	0	0	99	1	0	0
Mali	0	84	0	16	0	82	0	18
Malta	0	0	100	0	0	0	100	0
Mauritania	0	100	0	0	0	100	0	0
Mauritius	99	0	0	1	100	0	0	0
Mexico	51	4	0	45	24	5	0	71
Mongolia	99	0	0	1	100	0	0	0

TABLE 4. ***(Continued.)*** **Classification stability.**

Morocco	46	10	0	44	13	17	0	70
Mozambique	0	84	0	16	0	82	0	18
Myanmar	100	0	0	0	99	1	0	0
Namibia	50	2	0	48	24	2	0	74
Nepal	2	36	0	62	1	44	0	55
Netherlands	0	0	100	0	0	0	100	0
New Zealand	0	0	100	0	0	0	100	0
Nicaragua	46	10	0	44	15	12	0	73
Niger	0	84	0	16	0	82	0	18
Nigeria	0	100	0	0	0	100	0	0
Norway	0	0	100	0	0	0	100	0
Pakistan	0	43	0	57	0	51	0	49
Panama	83	1	0	16	68	0	0	32
Paraguay	46	10	0	44	13	16	0	71
Peru	51	4	0	45	24	5	0	71
Philippines	3	18	0	79	2	18	0	80
Poland	0	0	100	0	0	0	100	0
Portugal	1	0	99	0	2	0	95	3
Rep. of Korea	100	0	0	0	100	0	0	0
Romania	23	0	77	0	77	0	23	0
Rwanda	0	84	0	16	0	82	0	18
Saudi Arabia	17	58	0	25	7	69	0	24
Senegal	0	100	0	0	0	100	0	0
Sierra Leone	0	84	0	16	0	82	0	18
Singapore	99	0	0	1	100	0	0	0
South Africa	51	4	0	45	24	5	0	71
Spain	0	0	100	0	0	0	100	0
Sri Lanka	99	0	0	1	100	0	0	0
Sudan	0	84	0	16	0	82	0	18
Sweden	0	0	100	0	0	0	100	0
Switzerland	0	0	100	0	0	0	100	0
Syrian Arab Rep.	6	91	0	3	1	97	0	2
Taiwan	100	0	0	0	100	0	0	0
Thailand	100	0	0	0	100	0	0	0
Togo	0	100	0	0	0	100	0	0
Trinidad and Tobago	3	0	97	0	2	0	95	3
Tunisia	51	4	0	45	24	5	0	71
Turkey	83	1	0	16	68	0	0	32
Tanzania	0	43	0	57	0	51	0	49
Uganda	0	43	0	57	0	51	0	49
United Kingdom	0	0	100	0	0	0	100	0
United States of America	0	0	100	0	0	0	100	0
Uruguay	3	0	97	0	2	0	95	3
Venezuela	53	0	0	47	27	2	0	71
Viet Nam	100	0	0	0	100	0	0	0
Zambia	4	93	0	3	1	98	0	1
Zimbabwe	0	100	0	0	0	100	0	0

Percentage of classifications into cluster 1 (catching-up), 2 (poor), 3 (rich) or 4 (developing) based on different parameters. Part A shows the number of cases out of 100 models with different random initial partitions. Part B presents the number of cases out of 100 models with the number of concepts from 51 to 150. Other parameters as in the model with weights (see Table 2)

An alternative is also the analysis of the country-episodes, instead of countries, which accounts for the possibility of a within-country regime switch. All in all, the experiments point to the challenges associated with regime detection. The proposed method requires a researcher to make subjective decisions on the key parameters, and the decision itself has significant consequences on the obtained results. On the other hand, experimenting with a different set of parameters could lead to the detection of countries that escape a crisp classification.

VI. CONCLUSION

In the present paper, we introduced a novel procedure to identify economic growth regimes. To detect growth regimes, we proposed to adopt a fuzzy concept-based model. The model is based on a set of abstract concepts generated with the fuzzy c-means algorithm in multidimensional space. We run this model on a panel data set comprising a set of annual macroeconomic variables measured at the country level. To compare countries, the memberships for a given concept are summed across objects over all observations available. Then, in the last step, the Jaccard index is used to get a pairwise similarity measure, which serves as an input to a hierarchical clustering procedure.

The method allows to group countries that follow a similar growth regime, understood as a mix of growth performance and growth mechanisms.

We showcased the workings of the proposed method in three variants: the basic one, the variant with weights attached to input variables, and the variant with weights operating on country-episodes. Analyzing these variants, we showed that the proposed method could be adapted and tuned to a different research question. Importantly, each of the analyzed cases returned country clusters potentially informative and can be used as building blocks of follow-up studies.

Subsequently, we have investigated how robust our results are. Specifically, we tested the stability of classification into one of the discovered regimes under different numbers of concepts and different random initial partition matrices. We found many countries that are consistently classified as following one regime. At the same time, there are also countries whose classification changes together with the values of the parameters. To some extent, such a result is expected. Given the complexity of the economic growth phenomena, we can expect to find examples of economic development that cannot be reliably assigned to one of the detected regimes. Instead, we interpret these cases either as outliers or as countries whose development history shares the characteristics of several regimes. Thus, the procedure of our robustness analysis facilitates the identification of the cases that should be treated with additional caution.

REFERENCES

[1] F. C. Beylunioglu, M. E. Yazgan, and T. Stengos, "Detecting convergence clubs," *Macroeconomic Dyn.*, vol. 24, no. 3, pp. 629–669, Apr. 2020.

[2] D. Fiaschi, A. M. Lavezzi, and A. Parenti, "Deep and proximate determinants of the world income distribution," *Rev. Income Wealth*, vol. 66, pp. 677–710, Aug. 2019, doi: 10.1111/roiw.12418.

[3] O. Galor, *Unified Growth Theory*. Princeton, NJ, USA: Princeton Univ. Press, Dec. 2011, doi: 10.1515/9781400838868.

[4] P. Johnson and C. Papageorgiou, "What remains of cross-country convergence?" *J. Econ. Literature*, vol. 58, no. 1, pp. 129–175, Mar. 2020.

[5] A. L. Owen, J. Videras, and L. Davis, "Do all countries follow the same growth process?" *J. Econ. Growth*, vol. 14, no. 4, pp. 265–286, Dec. 2009.

[6] J. G. Brida, E. J. S. Carrera, and V. Segarra, "Clustering and regime dynamics for economic growth and income inequality," *Struct. Change Econ. Dyn.*, vol. 52, pp. 99–108, Mar. 2020.

[7] L. Pritchett, "Understanding patterns of economic growth: Searching for hills among plateaus, mountains, and plains," *World Bank Econ. Rev.*, vol. 14, no. 2, pp. 221–250, May 2000.

[8] A. Jastrzebska, G. Nápoles, Y. Salgueiro, and K. Vanhoof, "Evaluating time series similarity using concept-based models," *Knowl.-Based Syst.*, vol. 238, Feb. 2022, Art. no. 107811. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950705121010108

[9] J. Bartak and A. Jastrzebska, "Mining patterns of transitional growth using multivariate concept-based models," *Qual. Quantity*, vol. 56, no. 6, pp. 4395–4419, Dec. 2022.

[10] W. J. Baumol, "Productivity growth, convergence, and welfare: What the long-run data show," *Amer. Econ. Rev.*, pp. 1072–1085, 1986.

[11] S. N. Durlauf and P. A. Johnson, "Multiple regimes and cross-country growth behaviour," *J. Appl. Econ.*, vol. 10, no. 4, pp. 365–384, Oct. 1995, doi: 10.1002/jae.3950100404.
[12] C. M. Tan, "No one true path: Uncovering the interplay between geography, institutions, and fractionalization in economic development," *J. Appl. Econ.*, vol. 25, no. 7, pp. 1100–1127, Jun. 2009, doi: 10.1002/jae.1102.
[13] G. Di Vaio and K. Enflo, "Did globalization drive convergence? Identifying cross-country growth regimes in the long run," *Eur. Econ. Rev.*, vol. 55, no. 6, pp. 832–844, Aug. 2011.
[14] E. Flachaire, C. Garcia-Penalosa, and M. Konte, "Political versus economic institutions in the growth process," *SSRN Electronic J.*, vol. 42, no. 1, pp. 212–229, 2011, doi: 10.2139/ssrn.1831162.
[15] G. Liu, C.-C. Lee, and Y. Liu, "Growth path heterogeneity across provincial economies in China: The role of geography versus institutions," *Empirical Econ.*, vol. 59, no. 2, pp. 503–546, Aug. 2020, doi: 10.1007/s00181-019-01639-y.
[16] M. Jerzmanowski, "Empirics of hills, plateaus, mountains and plains: A Markov-switching approach to growth," *J. Develop. Econ.*, vol. 81, no. 2, pp. 357–385, Dec. 2006.
[17] M. Kerekes, "Growth miracles and failures in a Markov switching classification model of growth," *J. Develop. Econ.*, vol. 98, no. 2, pp. 167–177, Jul. 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0304387811000691
[18] B. Morier and V. K. Teles, "A time-varying Markov-switching model for economic growth," *Macroeconomic Dyn.*, vol. 20, no. 6, pp. 1550–1580, Sep. 2016.
[19] M. Peruzzi and A. Terzi, "Accelerating economic growth: The science beneath the art," *Econ. Model.*, vol. 103, Oct. 2021, Art. no. 105593. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0264999321001826
[20] A. Berg, J. D. Ostry, C. G. Tsangarides, and Y. Yakhshilikov, "Redistribution, inequality, and growth: New evidence," *J. Econ. Growth*, vol. 23, no. 3, pp. 259–305, Sep. 2018.
[21] A. Berg, J. D. Ostry, and J. Zettelmeyer, "What makes growth sustained?" *J. Develop. Econ.*, vol. 98, no. 2, pp. 149–166, Jul. 2012.
[22] J. Bezdek, *Pattern Recognition With Fuzzy Objective Function Algorithms*. New York, NY, USA: Springer, 1981.
[23] R. J. Hathaway and J. C. Bezdek, "Fuzzy C-means clustering of incomplete data," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 31, no. 5, pp. 735–744, Oct. 2001.
[24] R. C. Feenstra, R. Inklaar, and M. P. Timmer, "The next generation of the Penn world table," *Amer. Econ. Rev.*, vol. 105, no. 10, pp. 3150–3182, 2015.
[25] J. Bartak, L. Jablonski, and A. Jastrzebska, "Examining GDP growth and its volatility: An episodic approach," *Entropy*, vol. 23, no. 7, p. 890, Jul. 2021.

JAKUB BARTAK graduated from the Cracow University of Economics, in 2012. He received the Ph.D. degree in economics from the University of Rzeszów, in 2019.

In 2019, he joined as an Assistant Professor with the College of Management and Quality Sciences, Cracow University of Economics. His research interests include economics of long-term growth, income inequalities, and human capital.

AGNIESZKA JASTRZĘBSKA received the B.Sc. degree in information technology from the University of Derby, Derby, U.K., in 2009, the M.Sc. degree in computer engineering from the Rzeszow University of Technology, Rzeszów, Poland, in 2010, and the Ph.D. and D.Sc. degrees in computer science from the Warsaw University of Technology, Warsaw, Poland, in 2016 and 2021, respectively.

She is currently an Associate Professor with the Faculty of Mathematics and Information Science, Warsaw University of Technology, where she has been working since the beginning of her research career. Her research interests include machine learning, computational intelligence, and fuzzy modeling. She is an Associate Editor of the *Applied Soft Computing* journal.