ENDOGENOUS CATEGORIZATION OF THE HUMAN DEVELOPMENT*

Julio Abad-González† Universidad de León

Ricardo Martínez‡ Universidad de León

February 5, 2016

Abstract

The human development index (HDI) is one of the most well-known measures of welfare. We apply clustering techniques to endogenously determine how similar countries are with respect to the HDI, and into how many categories they can be classified. We find that, in contrast to the usual assumption in the United Nations' Human Development Reports, the number of categories is not fixed and has varied over time, from three in 1990 to four in 2014. We also find that the countries within each category differ from the United Nation's proposal.

Keywords: Human development, welfare, cluster analysis, agglomerative methods, grouping

JEL Classification: I30, D63

1 Introduction

Following the notion of functioning and capabilities proposed by Sen (1985), in 1990 the United Nations (UN) proposed a protocol to measure a country's overall degree of development using achievements in health, education, and per capita income as the keystones of welfare. The so-called *human development index* (HDI) was updated in 2010, coinciding with the 20th anniversary of the Human Development Report (HDR), to include several improvements.

Countries are ranked according to their HDI, which generates a classification comprising four categories: low, medium, high, and very high development. In the 2015 HDR, the HDI intervals corresponding to these four categories are $[0, 0.55)$, $[0.55, 0.7)$, $[0.7, 0.8)$, and $[0.8, 1]$. Beyond the

*This work was supported by Ministerio de Ciencia under Grants ECO2011-29355 and ECO2014-5376-P; Network MOMA under Grant ECO2014-57673-REDT, Junta de Andalucía under Grants SEJ5980 and SEJ4941; and Grupo PAIDI under Grant SEJ426.

†Departamento de Economía y Estadística, Universidad de León, León 24071, Spain. E-mail: julio.abad@unileon.es. Tel. (+34) 987 29 1753.

‡**Corresponding author.** Departamento de Economía y Estadística, Universidad de León, León 24071, Spain. E-mail: ricardo.martinez@unileon.es. Tel. (+34) 987 29 1911.

numerical value of the HDI, belonging to a particular category may be critical (international reputation and monetary transfers from other countries, for example). Although the choice of cutoff values for the intervals may seem reasonable, they are in fact arbitrary and exogenous. Besides, they need to be changed every few years to accommodate the evolution of human development. At this point, two questions arise. (1) Why should countries be classified into exactly four levels, especially when these are not symmetric? (2) Why were these particular cutoff values chosen?

To deal with the above issues, we apply clustering techniques to generate a classification that is endogenous and non-arbitrary, may vary over time, and does not require a predefined number of levels, since the levels are determined by the actual data.

The remainder of the paper is organized as follows. Section 2 introduces the clustering methodology. Section 3 presents the data and our main findings.

2 Methodology

One of the goals of cluster analysis is to determine natural groupings (called *clusters*) of observations.[1] Clustering methods are based on identifying a partition of observations such that observations within each cluster are as similar as possible, while observations between different clusters are as dissimilar as possible. These methods can be hierarchical or non-hierarchical. While we need to know in advance the number of clusters (which is not a trivial issue) for the latter, in the case of hierarchical analysis we do not require an a priori choice. Hierarchical clustering is typically implemented using agglomerative algorithms. At each stage of these algorithms there are several groups,[2] the two least dissimilar are merged (reducing the number of clusters by one), and the process is repeated. We apply the Ward algorithm discussed by Kaufman and Rousseeuw (1990) and Legendre and Legendre (2012), in which, at each stage, two clusters A and B are collapsed if their union minimizes the distance to any other cluster C according to

$$d(A \cup B, C) = \left[\frac{|A|+|C|}{|A|+|B|+|C|} d^2(A,C) + \frac{|B|+|C|}{|A|+|B|+|C|} d^2(B,C) - \frac{|C|}{|A|+|B|+|C|} d^2(A,B) \right]^{\frac{1}{2}}.$$

The results of such agglomeration algorithms are usually depicted by means of dendrograms showing the order in which clusters are formed and the distance spanned once they are combined.

3 Data and results

Our database consists of 144 countries and their HDI values in 1990 and 2014.[3] All the data were obtained from the UN Development Program and are available in Jahan (2015).

[1] Similar techniques have been applied to determine the formation of groups in other fields (see Lucotte (2015), for example.

[2] At be beginning of the process, all groups are singletons of just one observation.

[3] We consider these years for two reasons: the most recent data available are for 2014, and there are many missing values for dates before 1990.

Figures 1 and 2 show dendrograms for the HDI in 1990 and 2014, respectively; the horizontal axis represents the dissimilarity between countries, while vertical solid lines indicate the merger of two clusters. Clusters of countries that are very similar are linked at low distances, whereas clusters of countries that are very dissimilar are linked at high distances. It is evident that the number of clusters depends on the dissimilarity threshold chosen. According to Martínez (2011), human development has been more evenly distributed in recent years than in the 1990s; however, this overall result does not allow us to determine how similar or dissimilar countries are or into how many categories they should be classified. From Figure 1 it is natural to conclude that there were three different HDI levels in 1990; and countries with the poorest achievements (Group C) were significantly dissimilar to those in the top two levels (Groups A and B). In addition, countries in Group B are more similar to each other than to countries in the other two categories. Table 1 lists the nations within each level.

[Figure 1 about here.]

[Figure 2 about here.]

These findings are in contrast to the traditional HDR approach, which assumes that countries are classified into four groups for which the cutoff values are exogenously decided. We believe that the data are not consistent with such an assumption; even five groups would be more natural than just four. Figure 3 is a step graph that clarifies this point by plotting the number of clusters as a function of the distance or dissimilarity threshold. Thus, the longer the step, the more natural the number of clusters is. For 1990 (red solid line) we conclude that the optimal number of levels increases in the order two, three, five, four, ... Since the two-cluster case seems too drastic, three is the natural grouping (or even five), while the four-group choice should be dismissed.

[Figure 3 about here.]

After more than 20 years, the picture of the HDI in 2014 is rather different. Figure 2 shows that countries should be categorized into four levels (not three, as was the case in 1990). Besides, the similarity between the two groups with the greatest development (W and X) is almost equal to the similarity between the two groups with lowest development (Y and Z). Again, Figure 3 (dashed blue line) proves that four is the most reasonable number of clusters. Even though the number of categories coincides with the 2015 HDR, the countries that form them do not. Table 2 lists the nations within each level, where $\uparrow$ ($\downarrow$) indicates that a country is at a higher (lower) human development level compared to its category in the 2015 HDR. For instance, Kenya should be classified as having medium human development because it is more similar to nations in that category than to those with low human development.[4]

In recent years, the UN Development Program has proposed an alternative indicator to the HDI (called the *inequality-adjusted HDI*). This new index retains the essence of the HDI, but achievements in health, education, and income are adjusted according to their respective inequality (see

[4]This comparison is not so straightforward for 1990 since the number of levels is different.

Jahan (2015) for more details). It is worth mentioning that our conclusions are analogous for this adjusted index. Besides, our results do not differ qualitatively if we apply distances other than the usual choice in the agglomerative method for clustering.

In summary, we used agglomerative clustering techniques to classify countries into several categories according to their HDI, without setting in advance the number of categories there must be. We find that in 1990 there were three different human development categories, which increased to four in 2014. In comparison with the HDR, the application of the technique we propose results in a categorization where countries will be more similar within the levels we may obtain than within the levels imposed by the UN approach.

[Table 1 about here.]

[Table 2 about here.]

References

Jahan, S. (2015). *Human Development Report 2015*. United Nations.

Kaufman, L. and Rousseeuw, P. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley.

Legendre, P. and Legendre, L. (2012). *Numerical Ecology*. Elsevier.

Lucotte, Y. (2015). Euro area banking fragmentation in the aftermath of the crisis: a cluster analysis. *Applied Economics Letters*, 22:1046–1050.

Martínez, R. (2011). Inequality and the new human development index. *Applied Economics Letters*, 19:533–535.

Sen, A. (1985). *Commodities and Capabilities*. North-Holland.

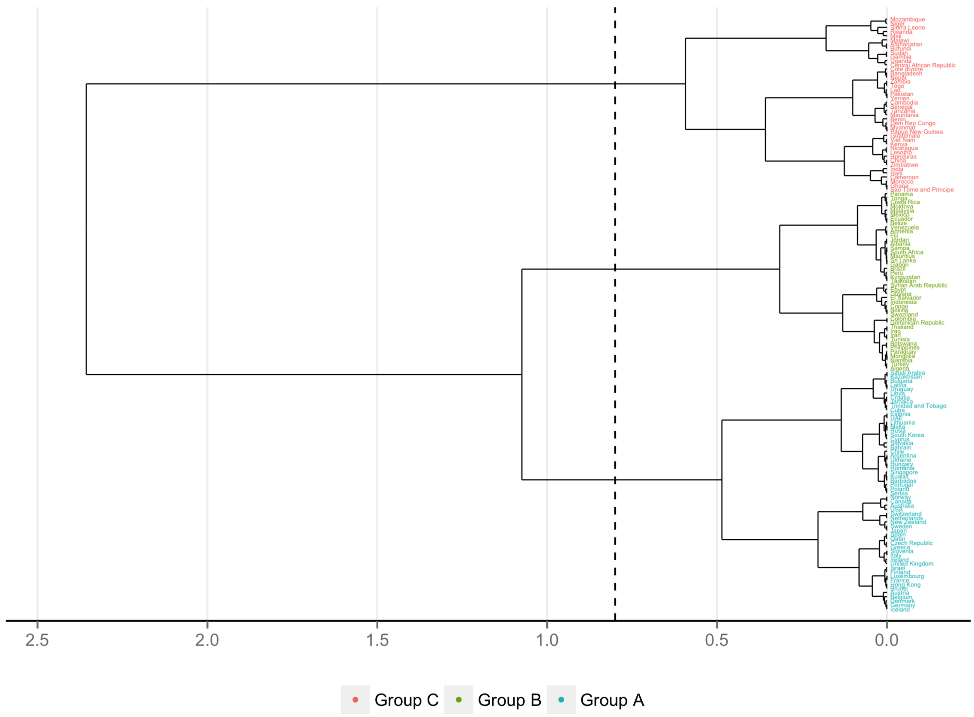

Figure 1: Dendrogram of the HDI in 1990.

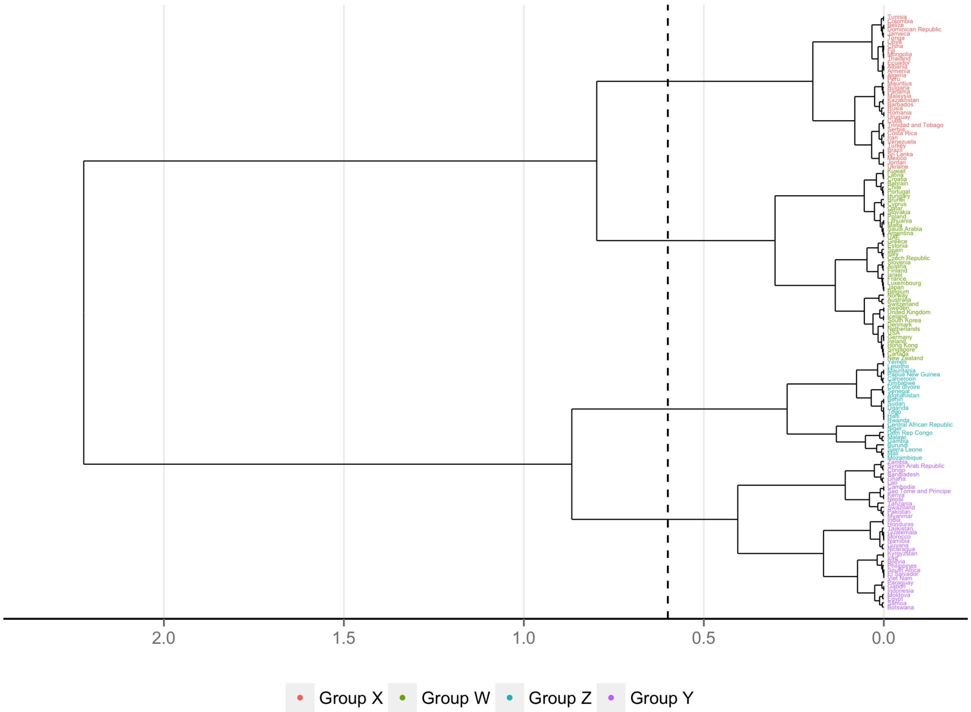

Figure 2: Dendrogram of the HDI in 2014.

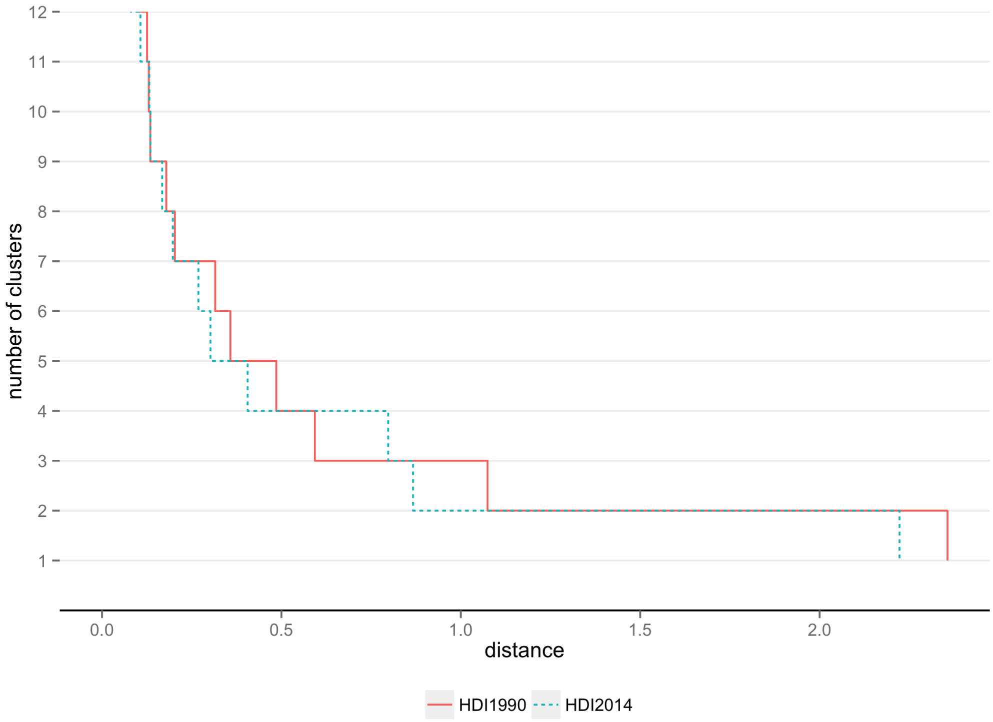

Figure 3: Number of clusters as a function of the distance.

Group A	Group B	Group C
Argentina	Albania	Afghanistan
Australia	Algeria	Bangladesh
Austria	Armenia	Benin
Bahrain	Belize	Burundi
Barbados	Bolivia	Cambodia
Belgium	Botswana	Cameroon
Brunei	Brazil	Central African Republic
Bulgaria	Colombia	China
Canada	Congo	Cote dIvoire
Chile	Costa Rica	Dem Rep Congo
Croatia	Dominican Republic	Gambia
Cuba	Ecuador	Ghana
Cyprus	Egypt	Guatemala
Czech Republic	El Salvador	Haiti
Denmark	Fiji	Honduras
Estonia	Gabon	India
Finland	Guyana	Kenya
France	Indonesia	Lao
Germany	Iran	Lesotho
Greece	Iraq	Malawi
Hong Kong	Jordan	Mali
Hungary	Kyrgyzstan	Mauritania
Iceland	Malaysia	Morocco
Ireland	Mauritius	Mozambique
Israel	Mexico	Myanmar
Italy	Moldova	Nepal
Jamaica	Mongolia	Nicaragua
Japan	Namibia	Niger
Kazakhstan	Panama	Pakistan
Kuwait	Paraguay	Papua New Guinea
Latvia	Peru	Rwanda
Libya	Philippines	Sao Tome and Principe
Lithuania	Samoa	Senegal
Luxembourg	South Africa	Sierra Leone
Malta	Sri Lanka	Sudan
Netherlands	Swaziland	Tanzania
New Zealand	Syrian Arab Republic	Togo
Norway	Tajikistan	Uganda
Poland	Thailand	Viet Nam
Portugal	Tonga	Yemen
Qatar	Tunisia	Zambia
Romania	Turkey	Zimbabwe
Rusia	Venezuela	
Saudi Arabia		
Serbia		
Singapore		
Slovakia		
Slovenia		
South Korea		
Spain		
Sweden		
Switzerland		
Trinidad and Tobago		
UAE		
USA		
Ukraine		
United Kingdom		
Uruguay		

Table 1: Level distribution of countries according to their HDI in 1990.

Group W	Group X	Group Y	Group Z
Argentina	Albania	Bangladesh	Afghanistan
Australia	Algeria	Bolivia	Benin
Austria	Armenia	Botswana	Burundi
Bahrain	Barbados	Cambodia	Cameroon
Belgium	Belize	Congo	Central African Republic
Brunei	Brazil	Egypt	Cote dIvoire
Canada	Bulgaria	El Salvador	Dem Rep Congo
Chile	China	Gabon	Gambia
Croatia	Colombia	Ghana	Haiti
Cyprus	Costa Rica	Guatemala	Lesotho
Czech Republic	Cuba	Guyana	Malawi
Denmark	Dominican Republic	Honduras	Mali
Estonia	Ecuador	India	Mauritania
Finland	Fiji	Indonesia	Mozambique
France	Iran	Iraq	Niger
Germany	Jamaica	Kenya$^{\uparrow}$	Papua New Guinea
Greece	Jordan	Kyrgyzstan	Rwanda
Hong Kong	Kazakhstan	Lao	Senegal
Hungary	Libya	Moldova	Sierra Leone
Iceland	Malaysia	Morocco	Sudan
Ireland	Mauritius	Myanmar$^{\uparrow}$	Togo
Israel	Mexico	Namibia	Uganda
Italy	Mongolia	Nepal$^{\uparrow}$	Yemen
Japan	Panama	Nicaragua	Zimbabwe
Kuwait	Peru	Pakistan$^{\uparrow}$	
Latvia	Romania	Paraguay	
Lithuania	Rusia	Philippines	
Luxembourg	Serbia	Samoa$^{\downarrow}$	
Malta	Sri Lanka	Sao Tome and Principe	
Netherlands	Thailand	South Africa	
New Zealand	Tonga	Swaziland$^{\uparrow}$	
Norway	Trinidad and Tobago	Syrian Arab Republic	
Poland	Tunisia	Tajikistan	
Portugal	Turkey	Tanzania$^{\uparrow}$	
Qatar	Ukraine	Viet Nam	
Saudi Arabia	Uruguay	Zambia	
Singapore	Venezuela		
Slovakia			
Slovenia			
South Korea			
Spain			
Sweden			
Switzerland			
UAE			
USA			
United Kingdom			

Table 2: Level distribution of countries according to their HDI in 2014.