Socio-Econ. Plann. Sci. Vol. 20, No. 4, pp. 237–241, 1986
Printed in Great Britain

0038-0121/86 $3.00 + 0.00
Pergamon Journals Ltd

CLASSIFICATION OF COUNTRIES FOR INTERNATIONAL DEVELOPMENT PLANNING USING CLUSTER ANALYSIS

JOHN C. S. TANG and P. A. SALVADOR
Division of Industrial Engineering and Management, Asian Institute of Technology, Bangkok 10501, Thailand

(*Received* 5 *October* 1985; *in revised form* 10 *January* 1986)

Abstract—This research study proposes an objective, systematic and multidimensional approach, which makes use of cluster analysis algorithms and complementary statistical techniques, to determine distinct classification or grouping of countries based on as many development indicators as thought relevant. The choice of development indicators to be included as variables is dictated by the interests of the development agency, which in this study is the Economic and Social Commission for Asia and the Pacific (ESCAP).

The results of the analysis suggest that there are five distinct groups of countries in the ESCAP region, rather than just three, known as developed, developing and least developed countries in ESCAP and other U.N. agencies. In both classifications, the developed countries and the least developed countries form distinct groups. However, the classification derived by cluster analysis refines the broad middle group of developing countries further into three; one group already closely similar to the developed group in many aspects, another group relatively nearer to the least developed group, and one group lying right in the middle of the middle layer. The analysis also directs attentions to possible misconceptions that could result from an assessment of developing countries based on just a few development indicators.

INTRODUCTION

During the past few decades, there has been a conscious effort to improve the living conditions in underdeveloped countries, to eradicate poverty and its concomittant ills: malnutrition, illiteracy, inadequate health services, among others. Many international organizations have sprung up to carry on the task of combatting the economic and social problems that plague many parts of the world, affecting hundreds of millions of people. Chief among such organisations is the United Nations (U.N.) and its affiliated specialized agencies and regional commissions such as the Economic and Social Commission for Asia and the Pacific (ESCAP). Even though significant and changes have been attained during the past 25 yr (see various issues of *World Development Report*). progress has been uneven and there are still million of people "... living in the very margin of existence with inadequate food, shelter, education, and health care. For many of them, there has been little improvement in the standard of living and for some, there may have been a deterioration..." [1(1978)].

The size and significance of the development tasks that remain to be done on an international scale call for objective assessment and systematic planning by concerned international organizations. It is in this area that this study hopes to contribute.

OBJECTIVE

The objective of this study is to develop and demonstrate a quantitative multidimensional approach to the classification of countries that can be used as a basis for development planning at the international level. Specifically, our goal is to group countries in the Asia–Pacific region by the use of cluster analysis and other complementary quantitative techniques in order to aid ESCAP in identifying the group or groups of countries that should be given high priority in development assistance.

ESCAP AND CURRENT CLASSIFICATION SYSTEM USED

The Economic and Social Commission for Asia and the Pacific (ESCAP) is one of the U.N. regional commisions with the task of coordinating and promoting the economic and social development of the countries of the Asia–Pacific region. It has a total membership of 44 countries.

Founded in 1947, ESCAP serves as a forum of Asian–Pacific governments where regional problems and international action plans for development are discussed and formulated. It also acts as a coordinator of the U.N.'s various specialized agencies which are operating in the Asia–Pacific region.

Currently, ESCAP has 44 member countries, 37 of which are full members, while the remaining 7 are associate members. Associate members are colonies protectorate or trust territories, which are as yet not independent countries. Five of the full members, all developed countries are not within the ESCAP region. Member countries geographically located within the ESCAP region are labeled regional members and those outside as non-regional members. The members of ESCAP are as follows:

Regional Members

1. Afghanistan
2. Australia
3. Bangladesh
4. Bhutan

5. Brunei
6. Burma
7. China (PROC)
8. Democratic Kampuchea
9. Fiji
10. India
11. Indonesia
12. Iran
13. Japan
14. Lao Democratic Republic
15. Malaysia
16. Maldives
17. Mongolia
18. Nauru
19. Nepal
20. New Zealand
21. Pakistan
22. Papua New Guinea
23. Philippines
24. Republic of Korea (South)
25. Samoa
26. Singapore
27. Solomon Islands
28. Sri Lanka
29. Thailand
30. Tonga
31. Vanuato
32. Vietnam

Regional Associate Members

1. Cook Islands
2. Guam
3. Hong Kong
4. Kiribati
5. Niue
6. Trust Territory of the Pacific
7. Tuvalu

Non-Regional Members

1. France
2. Netherlands
3. Union of Soviet Socialist Republics
4. United Kindgom of Great Britain and Northern Ireland
5. United States of America

For the purpose of assessing developing needs in order to enhance a more efficient allocation of efforts and resources, the above countries are generally classified by stages of development as follows:

(a) Least Development Regional Countries which include Afghanistan, Bangladesh, Bhutan, Lao People's Democratic Republic, Maldives, Samoa.
(b) Developing Countries which include all countries not listed in (a) and (c).
(c) Developed Countries which include Australia, Japan, and New Zealand.

Current approaches to grouping countries for assessing development needs are (1) qualitative assessment based on the analyst's perception on some aspects of development, (2) ranking on a single indicator such as GNP per capita and deciding on cut off points, and (3) ranking on a single indicator as in preceding procedure and then modifying the grouping based on a qualitative assessment of some other aspects of development.

Although, these approaches, when done thoroughly, can yield reasonable classification, still, its main disadvantage is that there is always the risk of depending too much on just one or a few indicators which are easily obtainable or are generally known. The use of cluster analysis on the other hand, ensures an objective multidimensional approach whereby many important indicators can be included and simultaneously considered, so long as the variables relevant to the purpose are identified. Cluster analysis offers a more logical approach that forces the analyst, just as in other quantitative techniques, to identify the relevant variables. Also, cluster analysis, unlike other classification techniques, whether qualitative or quantitative, does not require prior assumptions as to the structure of the data, but rather, draws structure out of the given set of data, minimizing the effects of possible biased assumptions.

OVERVIEW OF THE ANALYTICAL APPROACH

(*a*) *Choice of data units*

Nineteen countries have been included in the analysis. Even though the original objective was to include all ESCAP countries, however, for most of the new ESCAP members the available data are not complete or comparable with those of the older ESCAP member countries. The sample is deemed sufficient for the purpose of this study since the total population of the included countries account for 96% of the total population of ESCAP countries. The choice of the number of countries is related to the variables included in that, if it is desired to include all ESCAP countries, considerably less variables or indicators with available data can be included, and this may distort the analysis.

(*b*) *The variables used*

The appropriate variables that should be used are those that indicate level of development so that it can be said that members of the same group, derived by cluster analysis, are at fairly similar level of development, with the degree of similarity also depending on other chosen criteria. Through the years, several indicators have been consistently used and widely accepted by international development agencies as measures of a country's development. The development indicators used here are also the ones used by ESCAP and other international development organizations such as the World Bank, the Asian Development Bank, and other U.N. agencies in assessing levels of development. These indicators are: Basic Economic Indicators (GNP per capita, GNP per average annual growth rate, average annual rate of inflation, energy per capita consumption); Health and Nutrition Indicators (population/physician, population/nursing person, daily calorie intake per capita, daily calorie intake as % of requirement); Education indicators (adult literacy rate, number enrolled in primary school, number enrolled in secondary school, number in higher eduation); Life Expectancy and Demography Indicators

(Life expectancy at birth, infant mortality rate, population growth rate, fertility rate); Participation and Status of Women and Youth Indicators (ratio of economically active male population to economically active female population, ratio of male literacy to female literacy, ratio of male to female enrollment in secondary education, % of economically active youth); Labor Force Indicators (% of population of working age, % of labor force in agriculture, % of labor force in industry, % of labor force in services, average annual growth rate of labor force); Macroeconomic Growth Indicators (gross domestic product growth rate, average annual growth rate of agriculture, average annual growth rate of manufacturing, average annual growth rate of private consumption, gross domestic investment growth rate).

(c) *Method of standardization of data*

In all stages of the analysis, standardization of input data is by means of standard deviation or variance since this standardization method transforms both the means (transformed to zero) and the variance (transformed to one) and hence assures uniformity and comparability of data.

This is carried out by computing the standard deviation of the variable and dividing the values of the objects for each variable. Hence, the standardized values are in terms of standard deviation. The effect is similar to the familiar Z-transformation, whereby the standardized values are zero mean and variance of one, once the difference between two objects is taken. The standard deviation can be computed as follows:

$$S_j = \left(\frac{\sum_{i=1}^{m} (X_{ij} - \bar{X}_j)^2}{m-1} \right)^{1/2}$$

and

$$X'_{ij} = X_{ij}/S_j$$

where

X_{ij} = value of object i on variable j

$i = 1, 2, \ldots, m$

$j = 1, 2, \ldots, n$

$$\bar{X}_j = \sum_{i=1}^{m} X_{ij}/m$$

X'_{ij} = standardized value of variable i for object j.

Thus, when comparing two objects j and k on variable j the difference $dik = |X'_{ij} - X'_{kj}|$ is in terms of standard deviations. To compare results, do the Z-transformation:

$$Z_{ij} = \frac{X_{ij} - \bar{X}_j}{S_j}$$

Then the difference between the Z-transforms of i and k is

$$|Z_{ij} - Z_{kj}| = \left| \frac{X_{ij} - \bar{X}_j}{S_j} - \frac{X_{kj} - \bar{X}_j}{S_j} \right|$$
$$= |X'_{ij} - X'_{kj}| = dik$$

Hence, the difference of the standardized values dik is equivalent to the difference of the Z-transforms whose variance is one and mean is zero.

(d) *Measures of similarity*

Euclidian distance has been used as the measure of similarity and subsequently, as a relative measure of socio-economic development, given a reference group of developed countries. This approach does not require assumptions as to what should be the ideal development indicators of developed countries but only requires that actual developed countries in the sample be identified beforehand. The distance measure is further taken as indicator of "typicalness" and dispersion of countries in the group. Group characteristics are inferred from the means, which in fact, are the cluster centroids used as reference points for measuring between-cluster distances. Euclidian distance is used because it is intuitively more appealing and can thus contribute to a more meaningful interpretation of results.

(e) *Clustering criteria and methods*

The quantitative grouping of ESCAP countries is done in two stages, using complementary methods of cluster analysis as suggested by some authors [2]. The two stages of analysis to be followed here can be briefly described as follows:

(1) First, find an initial solution to the problem using the hierarchical clustering method of centroid linkage. This step is done on the assumption that little is known of the structure of data (i.e. how many groups there are and to which groups do the countries belong).
(2) This initial solution can then be refined in the second stage of analysis by using the no-hierarchical K-means clustering method, which requires as initial input the probable number of groups and initial cluster membership, if possible, so as to efficiently "arrive" at the final solution.

In the first stage centroid clustering procedure is chosen because of its ability to form compact clusters and also for convenience in interpretation, the centroid being simply the vector whose elements are the means of the indicators for the countries in the cluster. In the second stage the K-mean clustering procedure is applied because of its ability to refine solution, i.e. adjusting groupings such that an object is finally assigned to a group whose centroid is nearest to it.

(f) *Computer algorithm used*

Both stages of analysis are implemented using BMDP statistical software because it is an efficient package for clustering algorithms. For centroid clustering, BMDP, follows the agglomerative algorithm while the K-means clustering, division algorithm is used to attain the required number of clusters. A good reference on those algorithms can be obtained from the book by Spath [3].

RESULTS AND ANALYSIS

Classification of ESCAP countries derived by cluster analysis is presented in Table 1. (For details see [4].) For comparison purposes, Table 1 also contains the current ESCAP classification of the countries selected for this study.

It can be observed from Table 1 that that classification derived by cluster analysis, in general, does not conflict with ESCAP's current classification but, in fact, refines and reinforces it. Whereas, the current ESCAP classification recognizes the extremes of Developed and least Developed countries, the classification by cluster analysis breaks down the broad middle layer of developing countries into the finer categories, identifying 'tendencies' of countries toward either extremes. For instance, Singapore and Hong Kong are grouped separately, being more similar to the developed countries than the rest, while countries like Burma and India belong to a group with relatively lower level of socio-economic development and close to the least developed countries. It is particularly noteworthy that Papua New Guinea is identified as belongings to group 4 countries, considering that it per capita income is relatively higher than those in group 4 and even higher than that of Indonesia and Thailand. Under World Bank classification which is also used by ESCAP, Papua New Guinea is on a relatively high level within the middle-income group. The multidimensional analysis conducted here, in fact, shows it to be closer to the least developed countries, becuase of relatively low level of social development particularly in Education, Health and Nutrition.

There are some other observations about constituents of groups and current perceptions about these countries. Korea, which is often perceived as closely similar to Singapore and Hong Kong of group 2, falls under group 3 of the derived classification along with Malaysia, Thailand, Indonesia and the Philippines. A closer pairwise comparison of the input indicators show that Korea indeed has significant differences with either Singapore and Hong Kong, which on the other hand, are very similar. The wide differences between Korea and the other two countries are in GNP per capita, where Korea's is only one-third of the Singapore–Hong Kong group, and is just about equal to Malaysia.-Korea's inflation rate is much higher than those of the two city-state. and also, its economy as indicated by the distribution of its labor force, is relatively more dependent on agriculture and industry, wheras, Singapore and Hong Kong have an insignificant agriculture sector and are more service oriented.

Burma had GNP per capita income in 1982, and in previous years, at the same level as those of the U.N.-identified least developed countries, and in fact, lower than that of Afghanistan, yet is is not in group 5. A closer look at the development indicators for Burma reveals that except for GNP per capita, its level of socio-economic development, particularly for health, education and women participation, are within the range of those in group 4, rather than in group 5.

CONCLUSION

The preceding comparison illustrate some of the advantages of the proposed approach to classification of countries for international development planning. They can be summarized as follows:

(a) Objectivity—The basis for classification are measurable indicators, which when appropri-

Table 1. ESCAP classification and classification by cluster analysis

Current escap classification	Classification derived by cluster analysis
A. Developed Countries	A. Developed Countries (Group 1)
Australia	Australia
Japan	Japan
New Zealand	New Zealand
B. Developing Countries	B. Developing Countries I (Group 2)
Burma	Singapore
China	Hong Kong
India	
Indonesia	C. Developing Countries II (Group 3)
Malaysia	
Pakistan	Indonesia
Papua New Guinea	Malaysia
Philippines	Philippines
Republic of Korea	Republic of Korea
Singapore	Thailand
Sri Lanka	
Thailand	D. Developing Countries III (Group 4)
Hong Kong	
	Burma
C. Least Developed	China
	India
Afghanistan	Pakistan
Bangladesh	Papua New Guinea
Nepal	
	E. Least Developed Countries (Group 5)
	Afghanistan
	Bangladesh
	Nepal

†Source [4].

ately chosen would lead to objective classification. Except for a fore-knowledge of which countries are to be considered as developed countries, no other assumption is made about the structure of the data. Thus, "naturally" associated groups come out of the analysis.

(b) Multidimensionality—The method uses many variables considered as indicators of development, preventing misclassification that may be more likely to occur if only one or a few indicators (e.g. GNP per capita) is used. This property complements the belief that development is a multidimensional process.

(c) Discrimination—The proposed method is able to identify finer groupings. For instance, the resulting classification revealed that the current grouping of developing countries by ESCAP further refined into more distinct groups.

The refined grouping and the group development indicators suggest that indeed the identified least developed countries (namely, Afghanistan, Bangladesh, Nepal) should be given the highest priority. However, the analysis of group indicators show that some other countries (in particular, those in group 4) deserve high priority as well in some aspects of development, although they do not belong to the special group of least developed countries. The most pressing development needs of these groups are summarized below.

(1) Health and nutrition in group 5 countries.
(2) Productive capacity and income generation by groups 4 and 5, whose incomes are negligible fraction of the developed group 1 and even group 2 countries.
(3) Life expectancy is very low for group 5, indicating great need for improved conditions for basic necessities and for health care.
(4) Literacy for group 5 is very low (at 19%) and hence, more focus should be given to setting up and maintaining a broad based education system.
(5) Participation of women and youth in economic activities should be promoted in both groups 4 and 5 in as much that here they also lag far behind groups 1–3.
(6) More assistance in economic planning should be programmed for both groups 4 and 5 in as much as the long term trends indicate low economic growth rates.

In terms of ESCAP's activities, the above identified needs of the groups can be operationalized by organizing training programs especially for these groups, or more participants from these countries should be taken into ESCAP's training programs related to above identified needs. For instance, the seminars on health and services regularly held by ESCAP should take in more participants from group 5 countries. Also in-depth studies to identify the main and particular causes of low-level of development in the areas or development aspects previously identified, should be conducted.

A preliminary analysis of group development indicators similar to one done earlier—but more extensive—can be conducted during the preparation of programs and budgets, so that focus and participation in specific programs do reflect the needs of target groups. This is important considering that programs and specific activities have limited funding, so that only a limited number of participants can be accommodated.

REFERENCES

1. The World Bank. *World Development Report.* (various issues). Oxford University Press (1978–1984).
2. M. R. Anderberg. *Cluster Analysis for Applications.* Academic Press, New York (1973).
3. H. Spath. *Cluster Analysis Algorithms For Data Reduction and Classification of Objects.* Ellis Horwood, London (1980).
4. P. A. Salvador. *Classification of Countries for International Development Planning: A Quantitative Multidimensional Approach.* Research Study, M.Eng., SSPR No. IE 84-2, AIT (1984).