KATHOLIEKE UNIVERSITEIT TE LEUVEN
CENTRUM VOOR ECONOMISCHE STUDIEN

THE CLASSIFICATION OF COUNTRIES BY
CLUSTER AND BY FACTOR ANALYSIS

Lodewijk BERLAGE
and
Dirk TERWEDUWE *

Research Paper in Economic Development, nr. 8

June 1985

* The authors are grateful to P. Van Rompuy, E. Schokkaert and A. Decoster for helpful comments.

Contents

1. Introduction
2. Classifications of International Agencies
 - 2.1. The Least Developed Countries
 - 2.2. The Newly Industrialized Countries
3. Data
4. The Classifications Based on Cluster Analysis
 - 4.1. The Method
 - 4.2. General Discussion of the Results
 - 4.3. The LLDC's and the NIC's
 - a. The Least Developed Countries
 - (1) Which Countries ?
 - (2) Characteristics
 - b. The Newly Industrialized Countries
5. The Classifications Based on Factor Analysis
 - 5.1. The Method
 - 5.2. General Discussion of the Results
 - 5.3. The LLDC's and the NIC's
 - a. The Least Developed Countries
 - b. The Newly Industrialized Countries
6. Conclusion

References

Appendices

1. Introduction

The international organizations (U.N.O., World Bank, O.E.C.D.) classify countries into categories based on different criteria. Commonly used categories are e.g. the Least Developed Countries (LLDC's), and the Newly Industrialized Countries (NIC's). Which countries are included in the respective categories depends on the criteria used. There is no need to say that the process of establishing criteria for a list of countries that will receive benefits or that will be excluded from special treatment, is subjected to political pressures.

The purpose of this paper is to investigate in how far the classifications of countries, used by international agencies, can be justified by country characteristics. Two questions will be addressed. First, which countries constitute a relatively homogeneous group ? And second, which are the most outstanding characteristics of such a group ?

The analysis uses two different techniques, being cluster analysis and factor analysis. The first is a modern statistical method of partitioning an observed sample population into relatively homogeneous classes, called clusters. Factor analysis is a branch of multivariate analysis designed to explain the correlations or covariances among a set of variables in terms of a limited number of unobservable, latent variables. Both techniques are more or less complementary : factor analysis is a more appropriate technique to determine which variables are dominant or dormant in the classification, but it gives less clear clustering results of the observations (countries) than the technique of cluster analysis. Both methods make it possible to be less arbitrary in the choice of the variables which serve as a basis for classifications and thus in the classification itself.

Section 2 of this paper describes the category of Least Developed Countries and the category of Newly Industrialized Countries.

Section 3 examines the problem of the criteria used to establish classi-

fications of countries. The data used in our analysis are also given (1). Section 4 presents the classifications based on cluster analysis, while those based on factor analysis are analysed in section 5. In both sections our results are compared with the official categories of LLDC's and NIC's. Finally, the respective classifications are compared to each other and some conclusions are made.

2. Classifications of International Agencies

2.1. The Least Developed Countries

In an attempt to mobilize resources in favour of the poorest and weakest nations, a category of "least developed countries" was created by the United Nations. To identify these countries, the United Nations Committee for Development Planning (CDP) used three criteria :

- (1) a per capita gross domestic product of less than 100 US dollar (of 1968) ;
- (2) a share of manufacturing in gross domestic product of less than 10 % ;
- (3) an adult literacy rate of less than 20 %.

Based on these criteria the CDP recommended a list of 25 countries that was approved by the General Assembly in 1971. Since then 12 countries were added to the list, so that 36 countries (2) are now identified as least developed by the General Assembly with a total population of 280 million.

These are (3) :

26 countries in Africa : *Benin, Botswana, *Burkina Faso, *Burundi, Cape Verde, *Central African Republic, *Chad, Comoros, Djibouti, Equatorial Guinea, *Ethiopia, Gambia, *Guinea, Guinea-Bissau, Lesotho, *Malawi,

-
- (1) For a short description of the technique of cluster analysis and factor analysis, we refer to appendix A.
 - (2) Sikkim, one of the original countries on the list, ceased to be considered as an independent political entity.
 - (3) Countries which are included in the statistical analysis, are denoted by an asterisk.

*Mali, *Niger, *Rwanda, Sao Tome and Principe, *Sierra Leone,
*Somalia, *Sudan, *Tanzania, *Togo and *Uganda ;

9 countries in Asia and the Pacific : *Afghanistan, *Bangladesh,
Bhutan, *Lao People's Democratic Republic, Maldives, *Nepal, Wes-
tern Samoa, *Yemen Arab Republic, and *Yemen People's Democratic
Republic ;

1 country in the Caribbean : *Haiti.

The CDP's pragmatism in the selection of the indicators and its arbitrariness in establishing demarcation lines can be criticized (1). Our study will deal with this problem further on.

2.2. The Newly Industrialized Countries

Another group of developing countries which is often identified in work on development strategy and on trade and aid policy, are the so-called "Newly Industrialized Countries" (NIC's). This term refers to the group of countries which have recently become prominent as exporters of manufactured products. The composition of this group is not strictly defined. We use the term as defined by the Organization for Economic Cooperation and Development. This means that the group consists of the following 11 countries, four in Asia (*Hong Kong, *Republic of Korea, *Singapore and Taiwan), three in Latin America (*Argentina, *Brazil and *Mexico), four in Europe (*Greece, *Portugal, *Spain and *Yugoslavia).

Many other classifications can be thought of depending on the selected criteria. However, this paper focuses on the two important groups of countries mentioned above.

3. Data

As mentioned in section 2 of this paper, the classifications of countries used by international organizations are based on a few criteria only. Their

(1) WEISS T. and JENNINGS A. (1983), pp. 337-357.

classifications can be criticized because of the arbitrariness in selecting indicators and in establishing their critical values. Use of the techniques of cluster and factor analysis make it possible to overcome, at least in part, these problems. First, the problem of selection is overcome by simultaneously including a relatively large number of variables in the analysis. Second, and this is a special advantage of cluster analysis, there is no need to establish critical values for the variables.

For these reasons we have used cluster and factor analysis to analyze the classification of countries. Our work was based on twenty variables. Their selection was based on the fact that they are indicators of different country characteristics and on the availability of fairly reliable data for a large number of countries (1). The variables are :

- Variables related to income

1. GNPC : Gross National Product (GNP) per capita in dollars, 1981 ;
2. ENERC : Energy consumption per capita (kilograms of coal equivalent), 1980 ;

- Variables related to the structure of production

3. IND : Share of industry in Gross Domestic Product (GDP), 1981 ;
4. LABAGR : Percentage of labor force in agriculture, 1980 ;

- Variables related to exports

5. PRCOM : Primary commodities as share of merchandise exports, 1980 ;
6. EXP : Exports as percentage of GDP, 1981 ;

- Variables related to human resources

7. POP : Population size, mid 1981 ;
8. GPOP : Average annual growth of population (percent), 1970-1981 ;

(1) Source :

Variable 9 : O.E.C.D., Development Co-operation Review, series, and O.E.C.D., National Accounts Statistics, series ;
 Variable 12 : I.M.F., International Financial Statistics, series ;
 Variables 19 and 20 : O.E.C.D., Geographical Distribution of Financial Flows to Developing Countries, 1983.
 All the other variables : World Bank, World Development Report, series.

- Variables related to growth

- 9. GSR : Short run growth GNP per capita 1970-1980, (average annual growth rate) ;
- 10. GLR : Long run growth GNP per capita 1960-1980, (average annual growth rate) ;
- 11. GFOOD : Growth of food production per capita, 1970-1980 ;
- 12. GEXP : Average annual growth of exports in US dollars (percent), 1960-1980 ;

- Variables related to health

- 13. LIFE : Life expectancy at birth (years), 1981 ;
- 14. PHYS : Population per physician, 1980 ;
- 15. CAL : Daily per capita calories supply as percentage of requirement, 1980 ;

- Variables related to education

- 16. ALIT : Adult literacy (age group 15 years and over), percent, 1980 ;
- 17. PRIMS : Number enrolled in primary school as percentage of age group, 1980 ;
- 18. SECS : Number enrolled in secondary school as percentage of age group, 1980 ;

- Variables related to financial flows to developing countries

- 19. ODAC : Total net official development assistance per capita, averaged 1979-1981 ;
- 20. TRC : Total net receipts (1) per capita, averaged 1979-1981.

Data on all the variables were collected for as many countries as possible. In this way 102 countries were included in the analysis. Of these 84 are LDC's and 18 are industrialized countries.

The techniques of cluster and factor analysis were applied to two different data sets. Data set 1 consists of variables 1 to 18 for all 102 countries. Data set two includes financial flow variables 19 and 20 but excludes population size and growth (variables 7 and 8). It is limited

(1) The total net resource flow for a recipient country ("total net receipts") is the sum of official development assistance (ODA) + other official flows (OOF) + net private sector flows.

to the 84 LDC's (1).

4. The Classifications Based on Cluster Analysis

4.1. The Method

Cluster analysis is the modern statistical method of partitioning an observed sample population into relatively homogeneous classes, to produce an operational classification. The objective is to sort observations into groups called clusters so that the degree of statistical association is high among members of the same group and low between members of different groups.

Statistical association is expressed in terms of the difference between the values of the variables of cluster members and the cluster means. The partitioning into clusters is performed such that the sum of such differences over all variables, countries and clusters is minimized.

Details on clustering algorithms are given in appendix A. The algorithm that was used admits relocation of the countries when the number of clusters is reduced, thus allowing poor initial partitions to be corrected at a later stage.

Apart from the classification of countries by itself, it is also interesting to analyse which variables are dominant in the formation of clusters, for this purpose cluster analysis is of limited value (2). It can

(1) The analysis is based on standardized data.

The standardized value of variable i for country j is calculated as :

$$s_{ij} = \frac{x_{ij} - \mu_i}{\sigma_i}$$

where x_{ij} = value of variable i for country j

$$\mu_i = \frac{1}{n} \sum_j x_{ij} = \text{sample mean of variable } i$$

σ_i = sample standard deviation of variable i

(2) Factor analysis is a more appropriate technique to determine the dominant variables in the results, but it gives less clear clustering results than the technique of cluster analysis ; see section 5.

only describe the variables by which each cluster is characterized. This is done by hypothesis testing. Two tests are performed.

Consider the population of all countries under investigation. Assume this population is with respect to each variable i normally distributed with mean μ_i and standard deviation σ_i . Assume further that our clusters are samples taken from this normal population with cluster mean $\bar{x}_i$ and cluster standard deviation s_i . Then we can define $z_i = (\bar{x}_i - \mu_i) / (\sigma_i / \sqrt{n})$, with n = the number of countries in the cluster. If for a particular variable a cluster has a high z -value, this distinguishes the variable from the population in general.

The other test is based on the modified chi-square statistic, which is defined as $C_i^2 = s_i^2 / \sigma_i^2$. If the C^2 -value for a variable is low, this means for this variable that the difference between the values within the cluster is small relative to that in the population. Such a variable indicates a common characteristic of the countries in a cluster.

We call "dominant" those variables which satisfy both tests at the 99 % confidence level (see figures 2 and 4) (1).

In the remainder of this paragraph first the results of the analysis for both data sets are shown in a schematic way. Then some clusters are considered in more detail in order to compare the results with the classifications of the international agencies mentioned in section 2.

4.2. General Discussion of the Results

Figures 1 and 2 show the cluster results for data set 1. The figures show only the partition for nine, six, three and two clusters. Figure 1 visualizes which countries cluster together at each stage of the analysis. The aggregation of clusters is indicated by full lines whereas the movements of individual countries from one cluster to another, are visualized by dotted lines. Figure 2 links up with figure 1 and describes the variables by which the clusters of figure 1 are characterized at the 99 % confidence level (dominant variables).

(1) Detailed results of this hypothesis testing can be obtained from the authors.

Figure 1 : Cluster Analysis Using Data Set 1 : Both Developed and Developing Countries (102), Variables number 1 to 18

Note : * = Least Developed Countries.
° = Newly Industrialized Countries.

Figure 2 : Cluster Analysis Using Data Set 1 : the Dominant Variables for the Clusters of Figure 1
(99 % confidence level)

Note : The (+) and (-) refer to a positive and negative z-value respectively. Notice that C^2 is always positive

An analogous presentation of the results for data set 2 is given in figures 3 and 4.

Some general comments can be made. When the population variable (POP) is included, China and India appear as a separate cluster till we have less than six clusters (see figure 1). If this variable is left out China and India form no longer a separate group and belong to different clusters (figure 3). This points out the overwhelming dominance of the population variable for the cluster consisting of China and India.

A similar remark can be made about the variables related to financial flows going to developing countries. When they are included (data set 2) Jordan and Israel make up a separate cluster till we have less than six clusters (figure 3). Panama joins that cluster at the stage of six clusters. Leaving out the variables related to financial flows (figure 1) is sufficient to classify those countries into different clusters. This points out that the financial flows going to Israel and Jordan, and to a lesser extent to Panama, are so extremely high that it has become the most dominant characteristic of those countries.

Some countries seem to have such specific characteristics that it takes until the last stages of the analysis before they cluster with any other country. Examples are Singapore, Saudi Arabia and Libya (see figures 1 and 3). Although these countries are not very similar to each other (they do not cluster at an earlier stage in the analysis) they appear together at the stage of 6 clusters. Their cluster is characterized by a high energy consumption per capita, a large share of industry in GDP, a high level of exports relative to GDP, and both exports and GNP per capita growing fast during the period 1960-1980.

Figures 2 and 4 show the dominant variables in the analysis, i.e. variables whose values are rather similar for the countries in the cluster, but quite different from the average value of all countries. In both data sets the dominant variables in general are the share of industry in GDP (IND), the share of labor force in agriculture (LABAGR), life expectancy

Figure 3 : Cluster Analysis Using Data Set 2 : Only the Developing Countries (84),
Variables number 1 to 6 and 9 to 20

Note : * = Least Developed Countries.
° = Newly Industrialized Countries.

Figure 4 : Cluster Analysis Using Data Set 2 : the Dominant Variables for the Clusters of Figure 3
(99 % confidence level)

Note : The (+) and (-) refer to a positive and negative z-value respectively. Notice that C^2 is always positive

(LIFE), adult literacy (ALIT) and secondary school enrollment (SECS). Also important are the level of GNP per capita (GNPC), energy consumption per capita (ENERC) and per capita calories supply (CAL).

On the other hand, the variables related to exports (PRCOM and EXP), to population (POP and GPOP), to growth (GSR, GLR, GFOOD and GEXP) and to financial flows to the developing countries (ODAC and TRC) are rather dormant in the results.

4.3. The LLDC's and the NIC's

In this section we focus again on the categories of countries which were described in section 2, being the Least Developed (LLDC's) and the Newly Industrialized (NIC's). The comparison of the UN and OECD classifications with the corresponding clusters of our analysis leads to some interesting speculations.

a. The Least Developed Countries

(1) Which Countries ?

Of the 36 officially recognized LLDC's 24 countries were included in the analysis, denoted by an asterisk in section 2 of the paper. In the first partition of each analysis (9 clusters) these countries appear more or less as a separate group, which remains unchanged at the stage of six clusters. If we consider this cluster as the category of LLDC's, the analysis gives interesting results, which are summarized in table I. Countries included in the LLDC cluster in the respective analyses are denoted by an X.

Table 1 : Cluster Analysis (at the stage of 9 clusters) : the Category
of LLDC's

<u>Countries</u>	<u>Data set 1</u>	<u>Data set 2</u>
<u>Recognized LLDC's</u>		
Afghanistan	x	x
Bangladesh	x	
Benin	x	
Burkina Faso	x	x
Burundi	x	x
Central African Rep.	x	x
Chad	x	x
Ethiopia	x	x
Guinea	x	x
Haiti		
Lao PDR		
Malawi	x	x
Mali	x	x
Nepal	x	x
Niger	x	x
Rwanda	x	x
Somalia	x	x
Sierra Leone	x	x
Sudan	x	
Tanzania		
Togo		
Uganda	x	x
Yemen Arab Rep.	x	
Yemen PDR		
<u>Not Recognized</u>		
Mozambique	x	x
Senegal	x	x
Mauritania	x	x

If the countries which belong to the LLDC cluster for at least one data set are considered as least developed, five countries should be scrapped from the list of LLDC's as approved by the General Assembly of the United Nations : Haiti, Lao PDR, Tanzania, Togo and Yemen PDR. On the other hand, three countries should be added : Mozambique, Mauritania and Senegal.

However, the reader ought to be cautioned against rash conclusions. One should be aware of the methodological aspects of cluster analysis in itself and of the very important problem of the choice of variables to classify countries. One should also be aware of the problem that some countries may have made enormous efforts to score better on certain variables. For example Tanzania and also Lao PDR, both very poor countries, have made serious efforts to improve the basic needs of their people (1). Deleting them from a list that may lead to benefits, would then be a punishment for their efforts and for the results they obtained.

Finally when the number of clusters is reduced to three, all LLDC's except one belong to the same cluster. Yemen PDR, the only exception, is not classified with the other LLDC's and should clearly not be included in this class of countries.

(2) Characteristics

The creation of a group of LLDC's was inspired by an attempt to mobilize resources in favour of the poorest and weakest nations. Although some characteristics of poverty are obvious, it is not so easy to reach consensus on the criteria to identify least developed countries. This problem was somewhat overcome here by taking into account a large number of social and economic indicators. However, it may be interesting to investigate which are the most specific characteristics of the LLDC's as a group.

If we use as a criterion that variables should be significant for the cluster of LLDC's in at least one data set, then the following characteristics emerge :

(1) This will be demonstrated in section 5.2.

1. related to income level : GNP per capita and energy consumption per capita ;
2. related to the structure of production : share of industry in GDP and share of labor force in agriculture ;
3. related to health : life expectancy at birth and daily per capita calories supply ;
4. related to education : adult literacy and number enrolled in secondary school ;
5. related to growth : long run growth of GNP per capita and growth of food production.

Notice that the three variables used by the United Nations Committee for Development Planning are included in the list. This probably explains the great similarity between the list of countries which are officially recognized as LLDC's and those which emerge as LLDC's in our cluster analysis. An argument could be made to add a health indicator (life expectancy at birth, or daily per capita calories supply) to identify least developed countries.

As to the problem of establishing critical values for the selected indicators cluster analysis is not informative, but it is obvious that the CDP's method remains weak because of the arbitrariness in determining the cut-off point for each indicator.

b. The Newly Industrialized Countries

Following the O.E.C.D. in our analysis ten countries were considered as "newly industrialized" : Argentina, Brazil, Greece, Hong Kong, Korea Rep., Mexico, Portugal, Singapore, Spain and Yugoslavia. Taiwan was left out for lack of data. Figures 1 and 3 show that these countries do not emerge as a separate cluster. At the stage of nine clusters, the Latin American NIC's and Singapore are scattered over different clusters. Even at the step of three clusters, the Latin American NIC's (and with data set 2 Portugal also) stay separate from the other ones. This points out that one ought to be careful in interpreting a classification based on one com-

mon characteristic, being that they are exporters of manufactured products. Treating these countries as a homogeneous group for all kinds of purposes is not justified.

Nonetheless, six of these countries cluster in both analyses and thus constitute a rather homogeneous group : Greece, Portugal, Spain, Yugoslavia, Korea Rep. and Hong Kong. Distinctive characteristics of these six countries are a high life expectancy at birth and a high adult literacy rate.

Singapore and the three Latin American NIC's do not belong to this group. As already mentioned Singapore is a special case and seems difficult to classify, at an early stage, in a well specified group. The Latin American NIC's are classified with a cluster consisting of middle income and some low income countries.

5. The Classifications Based on Factor Analysis

We now focus on the classification results of a factor analysis (1). Are the results similar to those of the cluster analysis ? Before giving an answer to that question, some remarks on methodology ought to be made.

5.1. The Method

Factor analysis decomposes the variance of a variable into several components based on its association with latent variables. These variables are not observable : they are hypothetical constructs, called factors, which are derived from the original variables. The major aim of factor analysis is to determine the coefficients that relate the observed variables to the common factors. These coefficients are called factor loa-

-
- (1) The precise method used here is a principal factor analysis using squared multiple correlations for the prior communality estimates. To derive a simpler factor structure of the original factor solution, an orthogonal transformation employing Kaiser's varimax criterion is used. For a short description of factor analysis, see appendix A.

dings and they play the same role in factor analysis as regression coefficients in regression analysis.

If irrelevant variables are included in a factor analysis, this may lead to spurious results or random associations. To minimize such adverse effects two criteria have been used.

First variables with low factor loadings either have a low affinity with any of the factors or they are closely associated with more than one factor. Because such variables make the interpretation of the factor analysis less meaningful, they were deleted (1).

Second in order to have appropriate data for the factor analysis, the partial correlations should be small compared to the ordinary correlations. Variables for which the partial correlations are large relative to the ordinary correlations, were also omitted to better define the common factors (2).

The choice of the number of factors is always more or less arbitrary. In this paper the number of factors to be rotated is determined by the criterion that at least 90 per cent of the common variance must be accounted for by the retained factors.

5.2. General Discussion of the Results

The results of the factor analysis for data set 1 are presented in table II.

-
- (1) Variables whose highest rotated factor loading was below 0.50, were omitted from that particular data set. This criterion has been recommended by ADELMAN, I. and MORRIS, C.T. (1971), pp. 144-145.
 - (2) In ideal conditions factor analysis implies that the partial correlations among the variables, removing the effects of the common factors, are all zero. This is because, as the common factors are removed, only unique factors remain, which are by definition uncorrelated (see appendix A). Kaiser's measure of sampling adequacy (MSA) summarizes for each variable how much smaller the partial correlations are than the original correlations. Variables which have MSA's less than 0.50 were deleted. This criterion is recommended by the STATISTICAL ANALYSIS SYSTEM INSTITUTE (1982), p. 327.

Table II : Rotated Factor Loadings for Data Set 1 (All Countries, Variables 1 to 18).

Variables	Rotated factor loadings			Communality
	F ₁	F ₂	F ₃	
GNPC	.87	.12	.28	.85
ENERC	.83	.11	.30	.80
SECS	.72	.54	.27	.87
CAL	.62	.44	.37	.71
PRCOM	-.65	-.29	-.10	.52
LABAGR	-.67	-.57	-.31	.86
GPOP	-.77	-.20	.07	.64
PRIMS	.14	.78	.21	.68
ALIT	.57	.73	.09	.87
LIFE	.65	.68	.22	.94
PHYS	-.30	-.68	-.32	.65
GLR	.11	.29	.77	.70
IND	.23	.38	.66	.64
GSR	.10	.27	.66	.52
EXP	.12	-.02	.65	.43
GEXP	.14	.07	.60	.39
Sum of squared factor loadings	4.78	3.33	2.94	11.05

As was stated before data set 1 consists of all 102 countries and variables 1 to 18. Two variables were omitted : the population size (POP) because of too high partial correlations, and the growth of food production (GFOOD) because of a too low maximal loading. The percentage of common variance explained by the factors is 95.2 per cent (1).

(1) The over-all MSA is .90, ranging from .77 for GSR (short run growth of GNP per capita) to .97 for CAL (daily calories supply).

As explained, the factor loadings indicate the net correlation between each factor and the observed variables. The squared factor loadings represent the proportion of the total unit variance of variable i which is explained by factor p , after allowing for the contributions of the other factors. In table II it can be seen that $(.87)^2$ or 76 per cent of intercountry variations in per capita GNP is explained by factor F_1 , an additional 1.5 per cent ($= 0.12^2$) by factor F_2 , etc. The communality or the sum of the squared factor loadings, is an estimation of the total unit variance explained by all the common factors taken together, being, for example, 85 % for per capita GNP.

Table II provides the basis to group the variables into common factors. Each variable is assigned to the factor in which it has the highest loading. Once variables are assigned to the common factors, the factors must be "identified".

It may be observed that there exists a high correlation between factor F_1 and per capita GNP (.87). Factor F_1 may be interpreted as the "level of a highly developed modern society", characterized by a high per capita GNP (GNPC), which results in a high consumption of energy (ENERC) and calories (CAL), and a small traditional sector (negative coefficients for PRCOM and LABAGR). The negative coefficient for population growth (GPOP) is in accordance with the demographic view that a higher income leads to a lower birth rate. The high loading of secondary school enrollment (SECS) in factor F_1 (instead of factor F_2) may be unexpected at first sight, but an explanation for this may be that enrollment in secondary school is much more determined by economic forces than by social motivations.

The second factor may appropriately be called "basic needs satisfaction", with positive coefficients for primary school enrollment (PRIMS), adult literacy (ALIT), life expectancy (LIFE) and a negative coefficient for population per physician (PHYS).

Factor F_3 may be called "industrial export-led growth", with significant coefficients for long-run (GLR) and short-run (GSR) growth of per capita GNP, for the share of industry in GDP (IND) and for the export-related

variables i.e. the share of exports in GDP (EXP) and the growth rate of exports (GEXP).

We now focus on the classification of countries. For this purpose we have used the scores of individual countries on each factor. Such factor scores are formed as a linear combination of the variables using the factor loadings of the variables in each factor as weights, after standardizing all values. For each country we have three factor scores, one for each factor. These scores are weighted by the ratio of the sum of squared factor loadings for each factor to the sum of the communalities (see table II ; e.g. for F_1 $4.78 / 11.05 = 0.43$) and then added up. The weighted averages of the country scores for the three retained factors are given in table III.

A similar factor analysis was performed for data set 2, consisting of 84 (developing) countries only. This data set includes the variables related to the financial flows going to the developing countries and excludes the variables related to population. Two variables were dropped because of low maximum factor loadings : growth of food production (GFOOD) and the share of primary commodities in exports (PRCOM). The percentage of common variance explained by the three retained factors is 91.7 per cent (1). The results for data set 2 are shown in table IV.

-
- (1) The over-all MSA is .87, ranging from .45 for ODAC (official development assistance per capita) to .95 for PHYS (population per physician). In spite of the low MSA for ODAC, we retained the variable. Omitting ODAC led also to leaving out TRC (total receipts per capita), which has a MSA of .75, because of a low maximum loading for the last variable in that case. The addition of financial flows going to the developing countries is one of the main differences between data set 1 and data set 2.

Table III : The Classification of Countries by Weighted Average of Factor Scores Using Data Set 1 (All Countries, Variables number 1 to 18).

Lowest group ($x < -.5$)		Lower middle group ($-.5 < x < 0$)		Higher middle group ($0 < x < .5$)		Highest group ($x > .5$)	
*Chad	-1.00	Ivory Coast	-.49	Ecuador	.00	Israel	.55
*Uganda	-.83	Papua New Guinea	-.48	Turkey	.00	Libya	.55
Mozambique	-.82	*Haiti	-.46	Sri Lanka	.02	Saudi Arabia	.55
*Afghanistan	-.80	Cameroon	-.42	Congo PR	.05	New Zealand	.57
*Ethiopia	-.76	*Togo	-.40	China	.06	Trinidad-Tobago	.59
*Nepal	-.75	Kenya	-.40	Paraguay	.07	°Yugoslavia	.60
*Burkina Faso	-.75	India	-.39	Philippines	.09	°Korea Rep.	.62
*Niger	-.75	Zimbabwe	-.38	Colombia	.10	°Greece	.62
*Somalia	-.74	Liberia	-.38	Tunisia	.11	°Spain	.68
*Lao PDR	-.72	Burma	-.38	Syrian Arab Rep.	.11	Ireland	.68
*Burundi	-.71	Pakistan	-.37	Jamaica	.18	United Kingdom	.70
*Rwanda	-.71	Honduras	-.31	Malaysia	.18	Italy	.75
*Central Afr. Rep.	-.70	Bolivia	-.30	°Brazil	.18	Australia	.75
*Guinea	-.70	Nigeria	-.29	°Mexico	.18	°Hong Kong	.76
Senegal	-.67	Nicaragua	-.22	Chile	.18	Austria	.82
*Malawi	-.64	Morocco	-.22	Panama	.21	Finland	.82
Madagascar	-.63	El Salvador	-.22	Venezuela	.23	Denmark	.85
*Sudan	-.63	Indonesia	-.11	Costa Rica	.24	Switzerland	.85
*Sierra Leone	-.62	*Yemen PDR	-.07	Jordan	.27	France	.89
*Bangladesh	-.60	Dominican Rep.	-.07	°Argentina	.31	°Singapore	.91
*Mali	-.60	Thailand	-.06	Uruguay	.31	Netherlands	.93
Zaire	-.57	Peru	-.06	°Portugal	.45	Sweden	.93
*Tanzania	-.57	Egypt Arab Rep.	-.03			Japan	.94
*Yemen Arab Rep.	-.57	Algeria	-.02			Canada	.96
Mauritania	-.54					Germany Fed.	.96
*Benin	-.53					U.S.A.	.99
Ghana	-.52					Belgium	1.02
Zambia	-.52					Norway	1.06

Note : * = Least Developed Countries

° = Newly Industrialized Countries

Table IV : Rotated Factor Loadings for Data Set 2 (Only the Developing Countries, Variables 1 to 6 and 9 to 20).

Variables	Rotated factor loadings			Communality
	F ₁	F ₂	F ₃	
LIFE	.90	.29	.05	.89
ALIT	.89	.10	.00	.80
SECS	.79	.35	.21	.79
PRIMS	.78	.14	-.01	.63
CAL	.62	.49	-.02	.62
PHYS	-.71	-.27	-.05	.57
LABAGR	-.78	-.40	-.21	.81
GNPC	.30	.81	.06	.76
ENERC	.30	.79	.06	.71
EXP	.01	.75	.20	.60
IND	.44	.69	-.06	.68
GLR	.34	.66	.01	.55
GEXP	.13	.58	.21	.39
GSR	.31	.55	-.15	.43
TRC	.22	.31	.72	.66
ODAC	-.02	-.05	.68	.47
Sum of squared factor loadings	4.96	4.21	1.19	10.37

The three factors can be easily identified. Factor F₁ may be named "basic needs satisfaction", factor F₂ may be called "economic performance" and factor F₃ may be interpreted as "received financial flows from abroad". Notice that F₃, which is highly correlated with the total receipts per capita (TRC) and the official development assistance per capita (ODAC), is not associated with any other variable. This points out that the size of the financial flows going to different developing countries has nothing to do with their performance with respect to economic variables or basic needs indicators.

The classification of countries by weighted average of factor scores, based on data set 2, is given in table V.

Table V : The Classification of Countries by Weighted Average of Factor Scores Using Data Set 2 (Only the Developing Countries, Variables number 1 to 6 and 9 to 20)

Lowest group ($x < -.4$)		Middle group ($-.4 < x < .4$)		Highest group ($x > .4$)	
*Chad	- 1.11	India	- .37	°Brazil	.41
*Afghanistan	- .86	Pakistan	- .35	Malaysia	.43
*Uganda	- .86	Ivory Coast	- .35	Syrian Arab Rep.	.45
*Burkina Faso	- .86	Zambia	- .33	Congo PR	.46
*Ethiopia	- .85	Papua New Guinea	- .33	Jamaica	.46
*Niger	- .79	Burma	- .31	°Mexico	.48
*Burundi	- .79	Cameroon	- .30	Chile	.49
*Nepal	- .78	*Togo	- .24	Costa Rica	.52
Mozambique	- .76	Kenya	- .24	Uruguay	.57
*Central Afr. Rep.	- .70	Zimbabwe	- .17	°Argentina	.65
*Guinea	- .70	Liberia	- .15	Venezuela	.65
*Lao PDR	- .70	Nigeria	- .13	Panama	.71
*Rwanda	- .68	Honduras	- .12	°Portugal	.74
*Somalia	- .68	Bolivia	- .11	Jordan	.84
*Sierra Leone	- .65	El Salvador	- .10	°Yugoslavia	.87
*Malawi	- .65	Morocco	- .10	°Korea Rep.	.90
Senegal	- .64	Indonesia	.07	°Greece	.93
*Bangladesh	- .63	Nicaragua	.07	°Spain	1.04
*Mali	- .62	Thailand	.10	Trinidad-Tobago	1.12
*Sudan	- .60	Dominican Rep.	.13	Israel	1.15
*Yemen Arab Rep.	- .59	*Yemen PDR	.14	Saudi Arabia	1.19
Madagascar	- .55	Egypt Arab Rep.	.14	Libya	1.25
*Benin	- .49	Turkey	.19	°Hong Kong	1.25
Zaire	- .47	Sri Lanka	.20	°Singapore	1.64
*Haiti	- .46	Peru	.20		
*Tanzania	- .45	China	.22		
Ghana	- .45	Ecuador	.27		
Mauritania	- .42	Paraguay	.27		
		Algeria	.28		
		Philippines	.30		
		Tunisia	.35		
		Colombia	.36		

Note : * = Least Developed Countries.

° = Newly Industrialized Countries.

Till now we have ranked the countries according to a weighted average of the factor scores for the three retained factors. However, it is also possible to rank the countries according to their factor scores on each individual factor. We have done this for the three respective factors of data set 2. Detailed results can be found in appendix B.

With respect to basic needs, the countries that score very weak on this measure are almost all LLDC's. But also Mauritania, Senegal, Papua New Guinea, Ivory Coast and Mozambique rank very low. Surprisingly, Saudi Arabia has also a very low score. Another interesting observation which is in line with the results of the cluster analysis, is that some LLDC's perform rather well on this factor : Tanzania, Benin, Haiti, Lao PDR and Togo.

On the other hand some developing countries score highly on the factor of basic needs. Not surprisingly, these are Mediterranean and Latin American countries and Asian newly industrialized countries. However, a high score is also obtained by the Philippines and especially by poor Sri Lanka, confirming the good results obtained by this country from its basic needs strategy during the sixties and seventies.

The ranking of countries according to their performance on main economic aggregates as GNP, exports and their growth, gives a picture that is less clear. The LLDC's are still in the lower part of the ranking, but they are much more dispersed. Some non-LLDC's have low scores : Ghana, Mozambique, Nicaragua, Zaire and Madagascar. On the other side Malawi, Togo and especially Yemen PDR are LLDC's that perform very well.

Comparison of the performance of countries on basic needs and economic aggregates respectively, shows some remarkable differences. For example, Saudi Arabia scores best on the factor of economic performance, but it scores at the same time worst except one on the factor of basic needs fulfillment. Another interesting case is Sri Lanka : as a poor non-LLDC it

performs remarkably well concerning basic needs, but the economic performance is rather weak.

Finally, the countries can be ordered by the last factor (F_3), representing the received financial flows from abroad. Most financial flows are not going to the least developed countries. The countries which absorb most of the financial flows, in per capita terms, are Jordan, Israel and Panama. This corresponds to the results of the cluster analysis, in which Jordan, Israel and Panama joined together in a separate cluster when the financial flows going to the developing countries were included in the analysis (see figure 3).

5.3. The LLDC's and the NIC's

a. The least developed countries

In table III the countries are grouped into four subsamples, according to a somewhat arbitrary criterion of demarcation lines for the weighted average of factor scores (1). As is apparent from the table, the group of countries with the lowest factor scores ($x < -0.5$) corresponds remarkably well with the group of officially recognized LLDC's. The group includes almost all official LLDC's, except Haiti, Togo and Yemen PDR. On the other hand, seven non-LLDC's seem to have a lot of similarities with the officially recognized LLDC's, being Mozambique, Senegal, Madagascar, Zaire, Mauritania, Ghana and Zambia.

For data set 2, the countries are analogously grouped into three subsamples in table V. If we consider the group with the lowest factor scores as the group of least developed countries, the same non-LLDC's except Zambia are a member of that category. On the other hand, two officially recognized LLDC's seem to have more similarities with the middle group, being Togo and Yemen PDR.

(1) Remember that cluster analysis does not need such an arbitrary demarcation line to cluster the countries.

b. The Newly Industrialized Countries

In data set 1 the industrial market economies were included. As can be seen from table III, some NIC's are classified with industrial countries. Yugoslavia, Greece, Spain and Korea Rep. form a core. Ranking even higher are Hong Kong and Singapore and ranking lower is Portugal. But also non-NIC's are ranked with industrial countries : Israel, Trinidad and Tobago, and oil exporters Libya and Saudi Arabia.

The three Latin American NIC's don't join this core group. Brazil, Mexico and Argentina have a score in the neighbourhood of countries like Jamaica, Malaysia, Chile, Panama, Costa Rica and Jordan.

A similar story can be told on the basis of data set 2 (see table V). Again, we have a core consisting of Yugoslavia, Greece, Spain and Korea Rep., with Portugal ranked lower and Hong Kong and Singapore higher. Again, the three Latin American NIC's are separated from the core group.

6. Conclusion

The classifications of countries by international agencies can be criticized because of the pragmatism in the selection of criteria and because of the arbitrariness in determining the critical values for each of them. Cluster and factor analysis have the advantage that many variables can be included in order to make the selection procedure more general. Besides, there is no need to establish an arbitrary demarcation line for each variable.

Nonetheless, for a correct interpretation of our results, the reader must be aware of the methodological aspects of these techniques, of the problem of deciding which variables to include in the analysis and of the data problems. Not only does the exclusion of certain countries for lack of data influence the results ; one must also take into account the possible unreliability of the data, especially in poor countries.

In this paper we compared the category of least developed countries, as established by the General Assembly of the United Nations, and the category of newly industrialized countries as defined by the O.E.C.D., with the country classifications obtained from cluster and factor analysis. With respect to the category of least developed countries, the results are summarized in table VI.

Table VI shows that the list of Least Developed Countries, as approved by the General Assembly doesn't differ very much from the list of those countries obtained from our cluster and factor analyses. Nonetheless, the results strongly indicate that Mauritania, Mozambique and Senegal are unjustly excluded from the list. In all our analyses those three countries appear together with the least developed countries. Three other countries are on the borderline : Ghana, Madagascar and Zaire appear together with the least developed countries for the factor analysis, but not for the cluster analysis.

The explanation for this is that Mozambique scores very low on both basic needs and economic performance. For Mauritania and Senegal the main reason seems to be the weak performance on basic needs indicators, while for Ghana, Madagascar and Zaire it is the low score for the factor of economic performance that dominates (see appendix B).

On the other hand, table VI suggests that Togo and Yemen PDR do have characteristics different from those of most LLDC's. In all our analyses those two countries are not classified with the other least developed countries. On the borderline we have three other countries : Haiti, Lao PDR and Tanzania, which do not cluster with the other least developed countries in the cluster analysis.

The explanation for this is that Togo scores relatively high on both economic and basic needs indicators. For Yemen PDR the main reason has to be sought in the good economic performance, while for Haiti, Lao PDR and Tanzania the reason is rather their relatively good performance with respect to basic needs fulfillment.

Table VI : Cluster Analysis, Factor Analysis and the LLDC's.

		+ LLDC's -	
Cluster Analysis	Data set 1	Mauritania Mozambique Senegal	Haiti Lao PDR Tanzania Togo Yemen PDR
	Data set 2	Mauritania Mozambique Senegal	Bangladesh Benin Haiti Lao PDR Sudan Tanzania Togo Yemen Arab. Rep. Yemen PDR
Factor Analysis	Data Set 1	Ghana Madagascar Mauritania Mozambique Senegal Zaire Zambia	Haiti Togo Yemen PDR
	Data Set 2	Ghana Madagascar Mauritania Mozambique Senegal Zaire	Togo Yemen PDR

Note : + means that the non-LLDC is a member of the group of least developed countries in the specific analysis, - means that the LLDC is not a member of that group.

It should be noted that it is easier to add a country to a list that leads to benefits than to remove one. Some countries like Tanzania and Sri Lanka have made enormous efforts to satisfy basic needs and therefore score high on the corresponding variables. Deleting Tanzania from the list and excluding Sri Lanka is then a punishment for the results they obtained.

With respect to the class of Newly Industrialized Countries, we found in the factor analysis a hard-core group consisting of Yugoslavia, Greece, Spain and Korea Republic. In the close neighbourhood to them there emerged three other NIC's : Portugal, Hong Kong and Singapore. This group with the exception of Singapore also appeared as a rather homogeneous group in the cluster analysis. The three Latin American NIC's (Argentina, Brazil and Mexico) do have characteristics different from those of the others. Treating the category of NIC's as a homogeneous group for all purposes is therefore not justified.

REFERENCES

- ADELMAN, I., MORRIS, C.T. (1971), Society, Politics and Economic Development : A Quantitative Approach, second ed. (The Johns Hopkins Press, Baltimore).
- ADELMAN, I., MORRIS, C.T. (1982), "Factor Analysis and Development : A Reply", Journal of Development Economics, Vol. 11, no 1, pp. 113-123.
- ANDERBERG, M. (1973), Cluster Analysis for Applications (Academic Press, New York).
- BERLAGE, L. (1983), "De Minst Ontwikkelde Landen", De Wereld Morgen, Vol. 18, no 5, pp. 93-96.
- BUMB, B. (1982), "Factor Analysis and Development : A Note", Journal of Development Economics, Vol. 11, no 1, pp. 109-112.
- BOUROCHE, J., SAPORTA, G. (1980), L'Analyse des Données (Presses Universitaires de France, Paris).
- DELLAPORTAS, G. (1983), "Classification of Nations as Developed and Less Developed : An Arrangement by Discriminant Analysis of Socioeconomic Data", The American Journal of Economics and Sociology, Vol. 42, no 2, pp. 153-166.
- EVERITT, B. (1974), Cluster Analysis (Heinemann Educational Books, London).
- HARMAN, H.H. (1967), Modern Factor Analysis, second ed. (University of Chicago Press, Chicago).
- HEWITT, A. (1982), "EEC Policies towards the Least Developed : an Analysis" in STEVENS, C. (ed.), EEC and the Third World : a Survey 2 : Hunger in the World (Hodder and Stoughton, London), pp. 125-136.
- HICKS, N. (1982), "Sector Priorities in Meeting Basic Needs : Some Statistical Evidence", World Development, Vol. 10, no 6, pp. 489-499.
- I.M.F., International Financial Statistics, series (IMF, Washington).
- LAWLEY, D., MAXWELL, A. (1971), Factor Analysis as a Statistical Method, second ed. (Butterworth, London).
- O.E.C.D. (1972), "Problems of the Least-Developed Countries", in O.E.C.D., Development Co-operation Review (O.E.C.D., Paris), pp. 103-123.
- O.E.C.D. (1982), "Developing Country Exports of Manufactured Products : the Experience of the 'Second-tier' Countries", in O.E.C.D., Development Co-operation Review (O.E.C.D., Paris), pp. 123-132.
- O.E.C.D. (1982), Geographical Distribution of Financial Flows to Developing Countries (O.E.C.D., Paris).

- O.E.C.D., Development Co-operation Review, series (O.E.C.D., Paris).
- O.E.C.D., National Accounts Statistics, series (O.E.C.D., Paris).
- SAS-INSTITUTE (1982), SAS User's Guide : Statistics (SAS-Institute, Cary NC).
- U.N.I.D.O. (1974), "The Least Developed and Land-Locked Countries", in U.N.I.D.O., Industrial Development Survey (U.N.I.D.O., Vienna), pp. 260-273.
- WEISS, T., JENNINGS, A. (1983), "What are the Least Developed Countries and What Benefits may result from the Paris Conference ?", World Development, Vol. 11, no 4, pp. 337-357.
- WISHART, D. (1975), Clustan : Fortran IV Programs for Cluster Analysis (University College, London).
- WORLD BANK, World Development Report, series (The World Bank, Washington).

APPENDIX A

In this appendix we will give a short introduction to the statistical methods used in the paper. For a more extensive treatise, one should consult the texts listed in the bibliography.

1. The technique of cluster analysis (1)

Cluster analysis is the modern statistical method of partitioning an observed sample population into relatively homogeneous classes, to produce an operational classification. The objective is to sort observations into groups called clusters so that the degree of statistical association is high among members of the same group and low between members of different groups.

Cluster analysis is a term used to cover several different techniques, the two main categories being the hierarchical and non-hierarchical types. One algorithm of both types is discussed, being WARD's method and procedure RELOCATION respectively, both based on the error sum of squares.

1.1. WARD's method

a. Description of the algorithm

Ward's method is possibly the best of the hierarchical clustering techniques. It is an agglomerative method which proceeds by a series of successive fusions of the M entities (countries) into groups.

The basic procedure starts with the computation of a distance matrix between the entities. The most common distance measure is Euclidean distance.

Ward proposes that at any stage of an analysis the loss of information which results from the grouping of entities into clusters can be measured by the Error Sum of Squares (ESS). The ESS is defined as the sum of the distances from each entity to the centroid of its parent cluster. At each

(1) This section is based upon :
 ANDERBERG, M. (1973).
 EVERITT, B. (1974).
 WISHART, D. (1975).

step in the analysis, union of every possible pair of clusters is considered and the two clusters whose fusion results in the minimum increase in the ESS are combined.

To state this in a formal way, we define the following variables :

x_{ijk} = score of the i th variable for the j th country in the k th cluster, with $i = 1 \dots n$

$$j = 1 \dots m_k$$

$$k = 1 \dots h$$

$$\bar{x}_{ik} = \frac{1}{m_k} \sum_{j=1}^{m_k} x_{ijk}$$

= mean of the i th variable for countries in the k th cluster ;

$$\begin{aligned} ESS_k &= \sum_{i=1}^n \sum_{j=1}^{m_k} (x_{ijk} - \bar{x}_{ik})^2 \\ &= \sum_{i=1}^n \sum_{j=1}^{m_k} x_{ijk}^2 - m_k \sum_{i=1}^n \bar{x}_{ik}^2 \end{aligned}$$

= error sum of squares for cluster k , or the sum of Euclidean distances from each country in cluster k to the mean vector of cluster k ;

$$ESS = \sum_{k=1}^h ESS_k$$

= total within group error sum of squares for the collection of clusters.

The algorithm can be developed now. To begin with, there are m countries, each a cluster by themselves. Consequently the membership and the mean of each cluster coincide so that $ESS = 0$.

The Ward objective is to find at each stage those two clusters whose merger gives the minimum increase in the ESS. Suppose clusters p and q are chosen to be merged and the resulting cluster is denoted as t . Then the increase in ESS is :

$$\Delta ESS = ESS_t - ESS_p - ESS_q$$

$$\begin{aligned}
 &= \left[\sum_{i=1}^n \sum_{j=1}^{m_t} x_{ijt}^2 - m_t \sum_{i=1}^n \bar{x}_{it}^2 \right] - \left[\sum_{i=1}^n \sum_{j=1}^{m_p} x_{ijp}^2 - m_p \sum_{i=1}^n \bar{x}_{ip}^2 \right] \\
 &\quad - \left[\sum_{i=1}^n \sum_{j=1}^{m_q} x_{ijq}^2 - m_q \sum_{i=1}^n \bar{x}_{iq}^2 \right] \\
 &= m_p \sum_{i=1}^n \bar{x}_{ip}^2 + m_q \sum_{i=1}^n \bar{x}_{iq}^2 - m_t \sum_{i=1}^n \bar{x}_{it}^2 \quad (1)
 \end{aligned}$$

$$\text{Since : } m_t \bar{x}_{it} = m_p \bar{x}_{ip} + m_q \bar{x}_{iq} \quad \text{and} \quad m_t = m_p + m_q \quad (2)$$

$$\text{we get : } m_t^2 \bar{x}_{it}^2 = m_p^2 \bar{x}_{ip}^2 + m_q^2 \bar{x}_{iq}^2 + 2m_p m_q \bar{x}_{ip} \bar{x}_{iq} \quad (3)$$

$$\text{And as : } 2\bar{x}_{ip} \bar{x}_{iq} = \bar{x}_{ip}^2 + \bar{x}_{iq}^2 - (\bar{x}_{ip} - \bar{x}_{iq})^2$$

we get by substituting this last equation into equation (3) :

$$\begin{aligned}
 m_t^2 \bar{x}_{it}^2 &= m_p (m_p + m_q) \bar{x}_{ip}^2 + m_q (m_p + m_q) \bar{x}_{iq}^2 - m_p m_q (\bar{x}_{ip} - \bar{x}_{iq})^2 \\
 &= m_p m_t \bar{x}_{ip}^2 + m_q m_t \bar{x}_{iq}^2 - m_p m_q (\bar{x}_{ip} - \bar{x}_{iq})^2
 \end{aligned}$$

or deviding both sides by m_t^2 :

$$\bar{x}_{it}^2 = \frac{m_p}{m_t} \bar{x}_{ip}^2 + \frac{m_q}{m_t} \bar{x}_{iq}^2 - \frac{m_p m_q}{m_t^2} (\bar{x}_{ip} - \bar{x}_{iq})^2 \quad (4)$$

By substituting eq. (4) into eq. (1), we get :

$$\Delta ESS = \frac{m_p m_q}{m_p + m_q} \sum_{i=1}^n (\bar{x}_{ip} - \bar{x}_{iq})^2 \quad (5)$$

Thus, the minimum increase in the ESS is proportional to the squared Euclidean distance between the centroids of the merged clusters. Note that this method involves weighting of the distance between centroids when computing the distances.

To define the complete algorithm we still need an equation to update the distance matrix. Letting t denote the result of merging clusters p and q , and letting r denote any other cluster, the increase in the ESS that would result from the potential merger of clusters r and t is :

$$\Delta ESS_{rt} = \frac{m_r m_t}{m_r + m_t} \sum_{i=1}^n (\bar{x}_{ir} - \bar{x}_{it})^2 \quad (5\text{-bis})$$

$$= \frac{m_r m_t}{m_r + m_t} \sum_{i=1}^n (\bar{x}_{ir}^2 + \bar{x}_{it}^2 - 2 \bar{x}_{ir} \bar{x}_{it}) \quad (6)$$

By substituting eq. (2) and eq. (4) into eq. (6), we get :

$$\begin{aligned} \Delta ESS_{rt} &= \frac{m_r m_t}{m_r + m_t} \sum_{i=1}^n \left[\bar{x}_{ir}^2 + \frac{m_p}{m_t} \bar{x}_{ip}^2 + \frac{m_q}{m_t} \bar{x}_{iq}^2 - \frac{m_p m_q}{m_t^2} (\bar{x}_{ip} - \bar{x}_{iq})^2 \right. \\ &\quad \left. - 2 \bar{x}_{ir} \left(\frac{m_p}{m_t} \bar{x}_{ip} + \frac{m_q}{m_t} \bar{x}_{iq} \right) \right] \\ &= \frac{m_r m_t}{m_r + m_t} \sum_{i=1}^n \left[\frac{m_p}{m_t} (\bar{x}_{ir} - \bar{x}_{ip})^2 + \frac{m_q}{m_t} (\bar{x}_{ir} - \bar{x}_{iq})^2 - \frac{m_p m_q}{m_t^2} (\bar{x}_{ip} - \bar{x}_{iq})^2 \right] \\ &= \frac{1}{m_r + m_t} \sum_{i=1}^n \left[m_r m_p (\bar{x}_{ir} - \bar{x}_{ip})^2 + m_r m_q (\bar{x}_{ir} - \bar{x}_{iq})^2 - \frac{m_r m_p m_q}{m_p + m_q} (\bar{x}_{ip} - \bar{x}_{iq})^2 \right] \\ \Delta ESS_{rt} &= \frac{1}{m_r + m_t} \left[(m_r + m_p) \Delta ESS_{rp} + (m_r + m_q) \Delta ESS_{rq} - m_r \Delta ESS_{pq} \right] \quad (7) \end{aligned}$$

Each country is considered as a single cluster, hence $ESS = 0$.

(2) The two clusters whose merger results in the smallest increment in the ESS are 3 and 4. We label the new cluster 3.

(3) Updating of distance matrix :

$$\Delta ESS_{13} = \frac{1}{1+2} (2 \times 52 + 2 \times 58 - 1 \times 2) = 72.6$$

$$\Delta ESS_{23} = \frac{1}{1+2} (2 \times 61 + 2 \times 61 - 1 \times 2) = 80.6$$

$$\Delta ESS_{53} = \frac{1}{1+2} (2 \times 10 + 2 \times 4 - 1 \times 2) = 8.6$$

Using the replacement relation (8), ESS becomes :

$$ESS = 0 + \frac{1}{2} \cdot 2 = 1$$

Stage 2

(1)

$$D = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 5 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 5 \end{matrix} & \begin{bmatrix} 0 & & & \\ 5 & 0 & & \\ 72.6 & 80.6 & 0 & \\ 90 & 89 & 8.6 & 0 \end{bmatrix} \end{matrix}$$

(2) The smallest entry being 5, we merge clusters 1 and 2, with the resulting cluster labeled 1.

(3) Updating of distance matrix :

$$\Delta ESS_{31} = \frac{1}{2+2} (3 \times 72.6 + 3 \times 80.6 - 2 \times 5) = 112.5$$

$$\Delta ESS_{51} = \frac{1}{1+2} (2 \times 90 + 2 \times 89 - 1 \times 5) = 117.6$$

$$\text{And } ESS = 1 + \frac{1}{2} \cdot 5 = 3.5$$

Stage 3

(1)

$$D = \begin{matrix} & \begin{matrix} 1 & 3 & 5 \end{matrix} \\ \begin{matrix} 1 \\ 3 \\ 5 \end{matrix} & \begin{bmatrix} 0 & & \\ 112.5 & 0 & \\ 117.6 & 8.6 & 0 \end{bmatrix} \end{matrix}$$

(2) Fuse clusters 3 and 5, and label the new cluster 3.

(3) Updating of distance matrix :

$$\Delta ESS_{13} = \frac{1}{2+3} (4 \times 112.5 + 3 \times 117.6 - 2 \times 8.6) = 157.1$$

$$\text{And } ESS = 3.5 + \frac{1}{2} 8.6 = 7.8$$

Stage 4

(1)

$$D = \begin{matrix} & \begin{matrix} 1 & 3 \end{matrix} \\ \begin{matrix} 1 \\ 3 \end{matrix} & \begin{bmatrix} 0 & \\ 157.1 & 0 \end{bmatrix} \end{matrix}$$

(2) Fuse clusters 1 and 3.

$$(3) ESS = 7.8 + \frac{1}{2} 157.1 = 86.4$$

These results may be summarized as a dendrogram which is shown in the following figure.

Figure : Ward's clustering dendrogram.

1.2. Procedure RELOCATION

Relocation is a clustering technique which produces a partition of the countries, but differs from the hierarchical techniques (e.g. Ward's method) in that it admits relocation of the countries, thus allowing poor initial partitions to be corrected at a later stage. This advantage of relocation takes place in an attempt to optimize some clustering criterion, in our case the ESS.

This iterative procedure starts with a classification of the population of m countries into k clusters. During one relocation scan each country is considered in turn and its distances with all k clusters are computed. Suppose that the distance between country x and its parent cluster p is $d(p,x)$ and the distance between x and any other cluster q is $d(q,x)$. Then, if $d(p,x)$ exceeds $d(q,x)$, the method moves x from cluster p to cluster q , and thus leading to a lower ESS.

The centroids of clusters p and q are recomputed to account for this change at the time the switch occurs. The population is repeatedly scanned until no countries are relocated during one full scan, when a local optimum solution for k clusters will have been obtained. Whether this classification is the global optimum cannot be precisely established since, instead of examining each possible partition it rather re-arranges existing partitions.

Next, the distances between all pairs of clusters are computed and those two clusters whose merger leads to the least increase in the ESS are fused, thereby reducing the classification to $(k-1)$ clusters. The relocation phase is repeated to obtain a local optimum for $(k-1)$ clusters. Then the fusion phase is repeated to yield $(k-2)$ clusters. In this way, local optimum solutions are obtained for all clusterings from k to the minimum number of clusters required.

1.3. Ward or Relocation ?

The Relocation technique has the advantage over Ward's method in that it

admits relocation of countries, thus allowing poor initial classifications to be corrected at a later stage. On the other hand, it is an iterative procedure, which has the disadvantage that it cannot assure that the obtained classification is the global optimum.

We have done several cluster analyses based on both techniques. Because the results of both techniques were only slightly different, we may safely say that the Relocation procedure obtained a global optimum. In this case, the Relocation technique becomes superior to the method of Ward. Thus we preferred the Relocation procedure in our study.

2. The technique of factor analysis

The other technique used in this paper is factor analysis. This is a helpful method to reduce the number of characteristics by which countries are described. The large number of indicators are clustered in a small number of "factors" that must be interpreted intuitively. This smaller number of factors is then used to classify countries.

More strictly, factor analysis is a branch of multivariate analysis designed to explain the correlations or covariances among a set of variables in terms of a limited number of unobservable, latent variables, called factors, consisting of clusters of the original variables. The basic factor analysis equation can be written as :

$$z_{ij} = a_{i1} F_{1j} + a_{i2} F_{2j} + \dots + a_{iq} F_{qj} + U_{ij}$$

with z_{ij} = the observed value of country j on variable i

F_{pj} = the value of country j on the common factor p

a_{ip} = regression coefficient of the common factor p for predicting variable i

U_{ij} = the value of country j on the unique factor i

In this expression each of the n variables is described linearly in terms of q ($< n$) common factors and a unique factor. The coefficients of the factors are referred to as "factor loadings". The estimation of the loadings of the common factors is the basic problem of factor analysis. The unique factors play the role of residuals.

In matrix form the above equations reduce to :

$$Z = AF + U$$

where Z is the $(n \times m)$ matrix of observed country characteristics, A is the $(n \times q)$ matrix of factor loadings, F is the $(q \times m)$ matrix of common factor scores, and U is the $(n \times m)$ matrix of unique factor scores. Two critical assumptions are made :

- (1) the unique factors are uncorrelated with each other
- (2) the unique factors are uncorrelated with the common factors.

In our analysis all variables are standardized to unit variance. This unit variance of a variable can be expressed in terms of the factors analogous to the equation above. Then a common factor is an unobservable, hypothetical variable that contributes to the variance of at least two of the observed variables. A unique factor is an unobservable, hypothetical variable that contributes to the variance of only one of the observed variables. We have one unique factor for each observed variable.

The squared factor loadings represent the relative contribution of each factor to the standardized variance of each variable i , what leads to the concept of the "communality" of a variable i . This is given by the sum of squares of its factor loadings and indicates the extent to which the common factors account for the total unit variance of the variable i .

Another possibility of factor analysis is the expression of the common factor in terms of the observed variables. In this way factor scores for each country can be determined. The factor scores represent estimates of the values of each unobserved common factor for each country. Thus factor scores can be used to rank the countries with respect to the q common factors (1).

-
- (1) For more details about the methodology of factor analysis we refer to :
HARMAN, H.H. (1967).
LAWLEY, D., MAXWELL, A. (1971).

APPENDIX B

Table B.1 : The Ranking of Countries by Scores on the First Factor of Data Set 2 : "Basic Needs Fulfillment".

*Burkina Faso	- 1.72	Pakistan	- .56	°Brazil	.72
Saudi Arabia	- 1.65	*Haiti	- .54	Jordan	.72
*Niger	- 1.58	Nigeria	- .51	Congo PR	.75
*Chad	- 1.55	*Benin	- .45	Trinidad & Tobago	.76
*Afghanistan	- 1.49	Madagascar	- .36	Nicaragua	.77
*Ethiopia	- 1.45	Zambia	- .31	China	.84
*Burundi	- 1.41	India	- .24	Peru	.87
Mauritania	- 1.39	Zaire	- .18	Venezuela	.89
*Malawi	- 1.36	Morocco	- .17	Paraguay	.95
*Mali	- 1.25	Kenya	- .11	°Mexico	.99
Senegal	- 1.23	Ghana	- .08	Philippines	1.00
*Guinea	- 1.22	*Tanzania	- .08	Colombia	1.10
*Yemen Arab Rep.	- 1.22	Libya	- .02	Sri Lanka	1.16
*Central African Rep.	- 1.10	Algeria	.04	Israel	1.16
*Sierra Leone	- 1.04	Bolivia	.08	°Korea Rep.	1.17
*Somalia	- 1.00	Indonesia	.08	°Hong Kong	1.25
*Nepal	- .94	°Singapore	.10	Panama	1.25
Papua New Guinea	- .90	Burma	.17	°Portugal	1.29
*Rwanda	- .88	Honduras	.19	°Yugoslavia	1.38
*Yemen PDR	- .86	Zimbabwe	.23	Costa Rica	1.40
Ivory Coast	- .85	Egypt Arab Rep.	.24	Chile	1.41
Mozambique	- .77	El Salvador	.42	Jamaica	1.42
*Bangladesh	- .72	Tunisia	.43	°Greece	1.43
*Uganda	- .70	Malaysia	.52	°Argentina	1.56
*Sudan	- .68	Turkey	.55	°Spain	1.57
Liberia	- .67	Thailand	.56	Uruguay	1.60
*Togo	- .64	Dominican Rep.	.60		
Cameroon	- .59	Syrian Arab Rep.	.67		
*Lao PDR	- .56	Ecuador	.71		

* = Least Developed Countries.

° = Newly Industrialized Countries.

Table B.2 : The Ranking of Countries by Scores on the Second Factor of Data Set 2 : "Economic Performance".

*Uganda	- 1.24	Uruguay	- .39	Syrian Arab Rep.	.07
Ghana	- 1.05	*Sierra Leone	- .38	*Malawi	.11
*Lao PDR	- .98	Costa Rica	- .34	*Togo	.11
*Chad	- .95	Kenya	- .34	Egypt Arab Rep.	.13
Mozambique	- .94	Philippines	- .30	°Mexico	.15
*Tanzania	- .91	*Afghanistan	- .29	Ivory Coast	.16
Nicaragua	- .89	*Ethiopia	- .29	Mauritania	.21
Zaire	- .87	Colombia	- .27	Indonesia	.27
Madagascar	- .85	Dominican Rep.	- .24	°Brazil	.33
Burma	- .80	*Guinea	- .21	°Portugal	.39
Sri Lanka	- .77	Senegal	- .21	Tunisia	.42
El Salvador	- .71	°Argentina	- .20	Nigeria	.45
Jamaica	- .70	Paraguay	- .20	Malaysia	.54
*Benin	- .69	*Burundi	- .19	Israel	.56
*Nepal	- .69	Thailand	- .19	Venezuela	.64
*Sudan	- .66	*Mali	- .16	Algeria	.67
*Bangladesh	- .65	*Burkina Faso	- .11	°Yugoslavia	.68
*Somalia	- .62	China	- .09	°Greece	.75
Zimbabwe	- .53	*Niger	- .08	°Spain	.77
Honduras	- .49	Pakistan	- .05	°Korea Rep.	.88
Peru	- .48	Ecuador	- .03	*Yemen PDR	1.29
*Rwanda	- .48	Papua New Guinea	- .03	°Hong Kong	1.36
Zambia	- .48	Turkey	- .02	Trinidad & Tobago	2.02
India	- .46	*Yemen Arab Rep.	.02	Libya	2.75
Panama	- .44	Cameroon	.02	°Singapore	3.53
*Central African Rep.	- .42	Morocco	.04	Saudi Arabia	5.17
Bolivia	- .41	Jordan	.05		
Chile	- .41	Congo PR	.06		
*Haiti	- .39	Liberia	.07		

* = Least Developed Countries.

° = Newly Industrialized Countries.

Table B.3 : The Ranking of Countries by Scores on the Third Factor of
Data set 2 : "Received Financial Flows from Abroad".

China	- 1.30	*Haiti	- .33	*Central African Rep.	.09
Saudi Arabia	- 1.30	Tunisia	- .32	*Niger	.09
Paraguay	- .91	°Spain	- .32	*Benin	.10
Thailand	- .84	*Ethiopia	- .32	*Burkina Faso	.12
Indonesia	- .81	*Burundi	- .30	Bolivia	.13
°Brazil	- .73	*Tanzania	- .30	*Yemen PDR	.17
Trinidad & Tobago	- .68	Morocco	- .30	Zambia	.18
Burma	- .66	Egypt Arab Rep.	- .26	Ghana	.22
°Yugoslavia	- .65	*Guinea	- .24	*Togo	.28
Nigeria	- .63	°Korea Rep.	- .24	*Chad	.29
°Greece	- .60	Zaire	- .23	Senegal	.34
Turkey	- .57	Madagascar	- .22	*Mali	.39
°Mexico	- .57	°Argentina	- .22	*Somalia	.54
Dominican Rep.	- .57	Chile	- .22	Nicaragua	.61
Zimbabwe	- .55	*Lao PDR	- .19	Jamaica	.64
Philippines	- .55	Cameroon	- .19	Congo P.R.	.65
Ecuador	- .53	Algeria	- .18	°Hong Kong	.75
Colombia	- .53	*Afghanistan	- .16	Syrian Arab Rep.	.86
India	- .52	*Bangladesh	- .13	Papua New Guinea	1.00
Pakistan	- .48	Peru	- .12	Libya	1.18
*Rwanda	- .48	*Yemen Arab Rep.	- .09	°Singapore	1.29
Venezuela	- .44	Costa Rica	- .08	Liberia	1.31
°Portugal	- .43	*Uganda	- .08	Mauritania	1.46
*Malawi	- .41	Honduras	- .04	Panama	2.62
Sri Lanka	- .39	Ivory Coast	- .04	Israel	3.26
*Nepal	- .39	Mozambique	- .03	Jordan	4.22
Malaysia	- .38	El Salvador	- .02		
Kenya	- .37	*Sudan	- .01		
Uruguay	- .35	*Sierra Leone	.03		

* = Least Developed Countries.

° = Newly Industrialized Countries.

RESEARCH PAPERS IN ECONOMIC DEVELOPMENT

1. BERLAGE, L., CLAEYE, E. (1980), A Comparison of Official Development Assistance by Industrialised and Oil Exporting Countries.
2. STANDAERT, S. (1981), Money Demand under Repressed Inflation : Some Econometric Evidence for Algeria.
3. WATTEYNE, A. (1981), The Economics of Foreign Aid : Changing Perceptions.
4. BERLAGE, L. (1981), Is Development Aid Going to Poor Countries ? An Analysis of Bilateral Aid Flows from DAC-Countries in the 1970's.
5. BAECK, L. (1981), Income Distribution : French speaking Sub Sahara Africa ; The Slow Trickle Down of the Distribution Paradigm.
6. MOSS, F. (1982), Export Instability and its Causes : A Case Study of Tunisia.
7. STANDAERT, S. (1983), The Foreign Exchange Constraint, Suppression of the Trade Deficit and the Shadow Price of Foreign Exchange in a Fix-price Economy.
8. BERLAGE, L., TERWEDUWE, D. (1985), The Classification of Countries by Cluster and by Factor Analysis.