# WEIGHTING OF INDEX NUMBERS IN MULTILATERAL INTERNATIONAL COMPARISONS

BY LÁSZLÓ DRECHSLER*

The article reviews the methods used in practice and/or proposed by various authors for compiling indices in multilateral international comparisons. The various procedures are examined in the light of the following requirements: characteristicity (i.e. the weights should be characteristic to the countries which are compared), unbiasedness, circularity, internal consistency and factor relations.

There is no perfect solution since characteristicity and circularity are always and unbiasedness and internal consistency often in conflict with each other. The indices which are best for bilateral purposes are not transitive and the basic problem of multilateral comparisons is to obtain circularity, without losing too much of the characteristicity of the bilateral comparisons. Different compromises between the two requirements are possible and this is first of all what distinguishes the various methods used in practice.

Two main types of solution are applied in the various international comparisons. The first is based on the inter-spatial Fisher's ideal formula (e.g. the Eltetö–Köves–Szulc method, the van Yzeren method, the "central country" solution); the second type uses some kind of average prices (e.g. the Geary–Khamis method).

In the author's view there is no best method in absolute terms. Every method has some weaknesses and which of these weaknesses is the easiest to accept depends to a large extent on the actual aims of the comparison and on various other circumstances.

It is not surprising that interest in methodological problems relating to inter-country comparisons is rapidly growing. International comparisons of aggregates such as gross domestic product, national income, household consumption or industrial production, which were sporadically made before the sixties, are more and more frequently being undertaken. Some international organizations are compiling inter-country indices on an almost regular basis; others are undertaking *ad hoc* projects (sometimes on a very large scale) in these fields; there are, furthermore a number of inter-country comparisons carried out by pairs of countries, by the statistical offices or similar institutes of single countries, or by individual research workers.[1]

Among the various methodological problems of inter-country comparisons, perhaps the greatest interest is shown in the question of the types of index to be used for these purposes. What is the most appropriate formula for international comparisons is not an easy question, even if only two countries are concerned; and all these problems become much more complicated in multilateral comparisons, i.e. when more than two countries are involved. The present article is

*The author is a member of the secretariat of the United Nations Economic Commission for Europe. The views represented here are his own and not necessarily those of the secretariat.

[1]A few recent examples: under the auspices of the Conference of European Statisticians, Austria–Poland (household consumption), Czechoslovakia–France, Czechoslovakia–Hungary, Austria–Hungary (all industrial production and productivity); under the auspices of the CMEA comparisons of national income and its main components between the member countries. The most important present undertaking in this field is the United Nations–Pennsylvania University joint project on gross domestic product comparison, in the first stage of which ten countries are participating from different regions of the world.

primarily concerned with problems relating to the weighting of indices in multilateral comparisons. Problems of bilateral comparisons will be considered only insofar as they are relevant to multilateral comparisons.

Recently a number of articles and other studies have been published dealing with these problems.[2] Comparing them with some earlier studies[3] and with practices applied in the various inter-country comparisons, one finds some striking differences. Of course, the theory of index numbers has always been a field in which differences in views existed. However, in the present case, it seems that many of the differences are due to the fact that some questions of principle were not sufficiently well clarified. Furthermore, it seems that the work on inter-country index numbers was not sufficiently co-ordinated; most of the authors worked completely independently from the others and did not have access to the literature already available in this field.

The purpose of the present article is to review the earlier studies, and to draw some general conclusions from them. No attempt will be made to propose "the best method". In my opinion, there is no best method in absolute terms. That there is no perfect method is already fairly generally recognized. Every method has some weaknesses, and which of these weaknesses is the easiest to accept depends to a large extent on the actual aims of the given comparison and on various other circumstances. Thus, the relatively best method in each case can be selected only in the light of the principal aims and various circumstances of the given comparison. The present article tries to assist in the selection by giving a systematic presentation of the properties of the different methods proposed by various authors. These properties can be determined only in the light of the various requirements which the indices should satisfy.

## I. The Requirements to be Met by the Indices

The study of the various requirements the indices should satisfy has a long history. Irving Fisher's "tests" are well known and quoted by most of the authors. There is very little one can add to what has been said about these requirements, but one can be more selective. I shall deal here only with those requirements which have some relevance to the distinction of the various methods proposed (e.g. I shall not mention here those tests which are satisfied by every proposed method). On the other hand, I shall include some requirements which, though well known and recognized, are generally not considered as tests.

[2]E.g., S. Kawakatsu, "International Average Prices and Comparisons of National Aggregate Production of Agriculture, *The Review of Income and Wealth*, June 1970; S. Khamis, "A New System of Index Numbers for National and International Purposes", *Journal of the Royal Statistical Society*, Series A, Volume 135, Part 1, 1972.

[3]E.g., M. Gilbert–I. B. Kravis, *An International Comparison of National Products and the Purchasing Power of Currencies*, OEEC, Paris 1954; J. van Yzeren, *Three Methods of Comparing the Purchasing Power of Currencies*, The Netherlands Central Bureau of Statistics, Statistical Studies No. 7, 1956; Ö. Eltetö-P. Köves, "One Index Computation Problem of International Comparisons" (in Hungarian) *Statisztikai Szemle*, 1964. 7; B. Szulc, "Indices for Multiregional Comparisons" (in Polish), *Przeglad Statystyczny*, 1964.3; V. Strnad–E. Yershov, "Some Mathematical Problems Arising in the International Comparison of Economic Indicators", *Czechoslovak Economic Papers*, No. 5, 1965.

1. "*Characteristicity*" *of the Weights*

It is very difficult to find an appropriate term for this requirement (any suggestion for a better term would be welcomed). In general, this requirement means that the weights used for any index computations should be characteristic of the given two countries.

This requirement corresponds to the well known requirement of the "up-to-dateness" of weights in the case of inter-temporal indices. When constant prices are used in quantity index computations, it is recommended that after a certain period (generally 5 to 10 years), the weight base should be changed, i.e. new constant prices should be introduced, because the old ones are obsolete. In a more general sense, we may say now they are to be changed because they are no longer characteristic.

Coming back to the inter-spatial indices, the characteristicity requirement is satisfied if in the computation of indices the weights of the given two countries are used. In a Netherlands–Belgium quantity comparison, for instance, this requirement is completely satisfied if Dutch prices, Belgian prices or average Dutch–Belgian prices are used as weights. Average EEC weights are not fully characteristic for a Netherlands–Belgium comparison, and average European weights even less. To use Indian weights in a Netherlands–Belgium comparison would be considered wrong by everybody just as if in an Indian–Pakistan comparison Dutch weights were used. In the latter cases, the weights would be very uncharacteristic; their use would amount to the same as if in the case of the computation of a 1971–1970 inter-temporal index 1920 (or 2020) prices were used.

It is not possible to determine in general what importance should be attached to this requirement. If there is a particular interest in the results of each bilateral comparison, one should give due attention to the characteristicity of the weights. If, however, the interest in the binary results is only secondary to a more central interest (e.g. central EEC interest) it may be better to be more lenient in respect of characteristicity in order to satisfy some other requirements. I shall come back to these questions in a later part of this paper.

2. *Unbiasedness*

Bias, unbiasedness—as elsewhere in the index literature—is used here in a specific sense. It is a well known phenomenon that in inter-spatial quantity indices the use of own prices gives lower quantity indices for a given country than the use of the partner country's prices; and that in inter-country price index computations the use of own quantities as weights results in a relatively lower price level for the given country than the use of the other country's weights. The reason for this phenomenon is the negative correlation between the quantity and price proportions: in general, what is relatively cheaper is produced (or consumed) in a relatively greater quantity and vice-versa. Since in general there is no reason to assume that one of the two indices is better than the other, one considers that the use of own prices has a downward bias and the use of the partner country's prices an upward bias in the quantity index computations, in the same way as in the context of inter-temporal quantity indices one speaks about an upward bias of the Laspeyres formula and a downward bias of the Paasche formula.

One has to recognize that it is somewhat arbitrary to speak of biases in these cases. This consideration is based on the tacit assumption that the right answer is somewhere between the two indices. Of course, this can never be proved; but—and this should be repeated—since in general there are no reasons to contest that the two indices are equally good (or equally imperfect) this tacit assumption has its justification. The criterion of bias or unbiasedness therefore seems to be very useful.

It seems that practically all authors respect the unbiasedness requirement and try to satisfy it.[4] However, as will be shown later, not all methods proposed by these authors satisfy it equally well.

3. *Circularity (transitivity)*

This requirement is well known as Fisher's circular test. Having three countries, $A$, $B$ and $C$, this requirement means that index $B/A$ multiplied by index $C/B$ should equal index $C/A$.

It is useful to note at this stage already that the characteristicity and circularity requirements are incompatible with one another. One cannot satisfy both requirements by one set of indices. This can be explained without any recourse to mathematics. To be characteristic requires that each bilateral comparison ignore the "outside world". However, the "outside world" is always something else from bilateral comparison to bilateral comparison; and if one uses different weights, i.e. different yardsticks in each bilateral comparison, one cannot expect the requirement of circularity to be met.

Nobody contests that circularity is an important requirement in international comparisons. Without circularity one cannot get a consistent picture of the ensemble of countries and sometimes even the order of countries cannot be determined. However, views differ as regards what is the best way to obtain circularity, and what sacrifices should be made for this purpose in respect of the other requirements.

The relative importance to be attached to the circularity requirement depends, of course, to a large extent on the principal aims of the comparison. In general, one may say, the more dominating the central interest is over the particular binary (characteristicity) interests, the greater is the importance to be attached to circularity. In intra-community comparisons (e.g. intra-CMEA comparisons, intra-EEC comparisons) circularity is indispensable. In other cases, however, where the central interest is vague, where the participating countries do not constitute any ensemble, circularity may be subordinated to other requirements. It may also happen that one becomes interested in multilateral comparisons only after a number of binary comparisons have been accomplished: thus in a number of cases when one accomplishes a binary comparison, one does not know yet whether there will be a multilateral comparison or not, and if so, which countries will participate in it.

[4]The only recent exception seems to be A. Maddison, who preferred to use U.S. prices for a number of U.S.–other country comparisons, because "U.S. price structure reflected normal market forces more clearly". See "Comparative Productivity Levels in the Developing Countries", *Banca Nazionale di Lavoro, Quarterly Review*, Rome, December 1967.

4. *Internal (Structural) Consistency*

This requirement is perhaps less known, but its growing importance is widely recognized. Many of the classics on index numbers do not mention this requirement at all, the main reason being perhaps the lack of interest in structural comparisons.

The internal consistency requirement can be formulated in different ways. Köves in his book on indices[5] speaks of "average test" which means that an index of an ensemble must not be higher than the highest and not lower than the lowest of the subindices (i.e. subgroup indices). One can also be stricter in the formulation of this requirement. One may postulate that the indices should be consistent with some value data in comparable prices at each level of aggregation. In other words, there should be no conflict between the quantity indices and the "constant price" values (or their percentage distributions) of the countries for which the comparison is made. An example may help to explain better what this requirement means. Let us assume that after the elimination of price differences one has found that food consumption amounts to 40 percent of total consumption in country *A*, and to 44 percent in country *B*. These percentages should be consistent with the food consumption and total consumption *B/A* quantity indices, i.e. the ratio of the food *B/A* quantity index to the total *B/A* quantity index should be also 1.1 (the same as the ratio of 44 to 40).

This requirement has been one of the greatest troublemakers in recent international comparisons. In this connexion one should know that most of the studies used in some way or other the Fisher formula, which is on bad terms with this internal consistency requirement. One cannot connect the Fisher indices with any sets of percentage distributions or absolute value figures. This is perhaps also one of the main reasons why the Fisher formula could not gain sufficient ground in the field of inter-temporal comparisons. In national accounting the need for comparison of structures between different years (and after elimination of the impact of price changes) very often arises. While in the case of Laspeyres type (constant price) quantity indices there is no problem in having percentage distributions which are consistent with the indices, this is not true with the Fisher indices.

It is not possible to give a general appraisal of the relative importance to be attached to the internal consistency requirement. This may vary to a large extent from comparison to comparison. In some comparisons this requirement can be completely neglected. This is the case, for instance, when no comparisons of structures are envisaged or when the publication of the results is limited to the overall indices (the sub-indices being considered as not sufficiently reliable). In other cases, however, the comparison of structures may be important or even of central interest. In these cases the internal consistency requirement should be given priority over other requirements.

6. *Factor Relations*

The factor reversal test, which relates to the requirement that the product of the quantity and price indices should be equal to the ratio of values, is one of the oldest tests applied. It has relatively little relevance in the context of the

[5]P. Köves, *Statistical Indices* (in Hungarian), Budapest, 1956.

present study, since all of the methods which will be compared satisfy this requirement. Nevertheless, it should be mentioned here, partly because it is dealt with in practically all studies, and partly because here and there reference is made to some methods which do not satisfy the factor test (e.g. the Edgeworth–Marshall formula).

There is one further reason why it is worthwhile to have a closer look at this requirement. It seems that the factor test, as interpreted today by a number of authors, is not the same as this test in its original version. In the original version, the requirement was that the index which expresses the differences in the levels of the *actual* prices multiplied by the index which expresses the differences in quantity levels should equal the value ratio, the tacit assumption being that only actual prices are used in the computation of the quantity indices. It turns out, however, that in more recent studies actual prices are not the only prices used for valuation purposes. In cases where goods and services are highly subsidized, it is considered that the actual price does not provide the best approximation of the relative importance of the given commodity, and imputed rather than actual prices (mostly "cost" prices) are used in the valuation of aggregates for the quantity index computation.

When this is done, a distinction should be made between (i) price indices which express the difference between the levels of the *actual* prices and (ii) price indices which express the difference between the levels of the prices *used for the computation.* Only the second price index together with the quantity index satisfies the factor reversal test.

This does not deny the importance of the factor test. Even in situations where not only actual prices are used for the valuation, the factor test has an important function. One should not forget here that in many cases quantity indices cannot be compiled directly but only by the deflation of value data by means of price indices. In other cases, quantity indices can be compiled directly as well as indirectly, but even in these cases preference should be given to the indirect method, since price indices are generally subject to smaller incompleteness (sampling) errors than quantity indices. In these instances, the price indices to be used for deflation (conversion) should, of course be indices of "computation prices", and not of actual prices. In this context, the meaning of the factor test is no longer the same as interpreted by the classics.

It is perhaps worthwhile making a distinction between (i) the economic factor test, i.e. the factor test as interpreted by the classics, and (ii) the technical factor test where the price index is an intermediate tool for the quantity index computation rather than an aim in itself. The technical factor test should always be satisfied; but one cannot be so strict in respect of the economic factor test.

The Austria–Poland comparison of levels of consumption, carried out under the auspices of the Conference of European Statisticians, provides a good illustration of this problem. In both countries, particularly in Poland, a number of items (medicines, housing, educational and health services, etc.) were valued at other than actual prices. Price indices used for conversion from schillings to zloties and vice-versa were, of course, "computation price" indices and not "actual price" indices. This is strongly emphasized in the report on this comparison, which draws the attention of readers to the fact that the zloty/schilling

price indices which appear in the report do not, for the above reasons, reflect the purchasing power ratio of the two countries.

## II. The Methods of Bilateral Comparisons

Before turning to multilateral comparisons proper, it is worthwhile devoting a little time to the field of binary comparisons. To do this is expedient not only because many of the multilateral problems are already present in the methods of binary comparisons, but also because—as experience has shown—many of the multilateral studies emerge from bilateral comparisons.[6]

As to the question of what is the best method of bilateral comparisons, views as well as practices do not differ much. In fact, in most cases the inter-spatial Fisher formula, i.e., the geometric average of own weighted and partner country's weighted indices was used and/or proposed. This method was used in all studies carried out under the auspices of the Conference of European Statisticians and—with some modification to which I shall come back—in the various OEEC studies. It is also the method applied in the present United Nations Product and Purchasing Power Comparison Project. The inter-spatial Fisher formula was proposed for bilateral comparisons by van Yzeren, Köves, Eltetö, Szulc, Ribakov, Krzeczkowska and many other authors.

Let us now examine, in the light of the above requirements, how good the Fisher formula is for bilateral comparisons. There is no doubt that the characteristicity requirement is satisfied. The Fisher formula is not subject to a bias in the sense that we used it in the preceding section. Since we are considering here bilateral comparisons only, the requirement of circularity does not arise. The Fisher formula also satisfies the factor reversal test. The only trouble is—and this has already been mentioned—the Fisher formula does not meet the internal consistency requirement.

One may get into trouble with internal consistency in various ways. One of the possible symptoms of this drawback is that there may be a breach of the "average test" (Köves' terminology). To my knowledge, it never occurred in practice that the Fisher index was higher than the highest or lower than the lowest of the sub-indices. This is not so surprising. While it is not difficult to construct artificial examples where the Fisher formula does not satisfy the average test, a closer examination reveals that it only fails in cases where a number of special factors occur jointly, and that the likelihood that this will happen is minimal.

On the other hand, lack of consistency between percentage distributions and indices is quite often found. In many studies not only absolute levels but also relative structures were compared, and as has already been noted, one cannot construct distribution percentages which are consistent with the Fisher formula. In such situations there is perhaps no better solution than what was done in the Austria–Poland study and the CMEA comparisons, namely, to compile the percentage distributions on the basis of one of the countries' prices (e.g. in the

[6]E.g., under the auspices of the Conference of European Statisticians, on the basis of the results of three bilateral productivity comparisons (Czechoslovakia–France, Austria–Hungary, Czechoslovakia–Hungary) carried out originally independently from each other, a quadrilateral comparison of labour productivity was prepared.

Austria–Poland study on the basis of Austrian schilling data), and to use these percentages together with the Fisher indices in the analysis of the results.

This is, of course, not a perfect solution. The conditions for conflict are present. For instance, it may happen that the $B/A$ quantity index of milk is higher than the $B/A$ quantity index of meat; nevertheless, meat has a higher relative proportion in country $B$, while milk a higher relative proportion in country $A$. In practice, however, the situation is not so bad. The extent to which conflicts arise is in most cases relatively small. For instance, the reader of the Austria–Poland study who is not already familiar with this type of methodological problem may even not have noticed the presence of this difficulty.

So much for the Fisher formula. In only a few cases have formulae other than Fisher's been used or proposed for inter-country comparisons. It has already been noted that the "classical" OEEC studies were also exceptional in a certain sense. In fact, in the Gilbert–Kravis and other OEEC studies, in many cases no geometric average was computed from the two basic ("Laspeyres" and "Paasche" type) indices, and these latter were used as final results of the comparison.

This attitude is perhaps a late after-effect of the great index debates of the twenties, when many who opposed the Fisher index accepted both the Laspeyres and Paasche formulae but objected to the compilation of any kind of average.

One can hardly criticize this method on theoretical grounds. However, from a practical point of view it is rather uncomfortable to use both indices as final results. It is perhaps not too bad to have two answers to a simple question. However, there are more involved tasks in the studies. For instance, if one wants to compare consumption differences with production differences and to do the same for different periods, then the duplication of the answers may be really annoying, and disturbing to the perspicacity. The OEEC authors also felt these disadvantages, and in places where the analysis became multidimensional,[7] they did use the average of the two basic indices.

It seems that apart from A. Maddison, referred to above, nobody proposes to use the weights of one of the participating countries only. The main disadvantage of using such a one-sided weighting is the bias. One can hardly expect that Maddison's argument (i.e. that the prices of one country may reflect truer market forces than the prices of the other country), together with the avoidance of internal consistency problems, will generally compensate for this important drawback.

Apparently, nobody proposes any more to compile averages other than geometric ones from the two basic indices. The arithmetic average which was proposed and/or used in some earlier studies would have one advantage but two important disadvantages as compared with the Fisher formula. It would more adequately meet the internal consistency requirement, but it would not satisfy the factor test, and what is perhaps the most important, the final results, i.e., the final $B/A$ index, would be different depending on which country was originally the numerator and which the denominator (in the inter-temporal index language it would not satisfy the time reversal test).

[7]See e.g., M. Gilbert and Associates, *Comparative National Products and Price Levels*, Paris 1958, page 36.

On the other hand, there are recent proposals and even practical experiences in using average weights for the computation of indices. To compile averages of the weights instead of averages of the indices is not a new idea in the index theory. The basic considerations are very similar: one wants to avoid bias, and average weights can be as neutral as averages of the indices.

The main reason why averaging of weights did not gain ground in earlier practices is presumably the fact that this method presents a number of practical difficulties. One of them is the apparent *circulus vitiosus* that (i) for obtaining the right average prices as weights of the quantity indices one would need first to have price indices and (ii) for obtaining the right average quantities as weights of the price indices it is desirable to have first quantity indices.

Not so long ago S. H. Khamis, utilizing the idea of an earlier study of R. C. Geary,[8] presented a method which overcomes this difficulty. He solves this *circulus vitiosus* by a system of linear equations. This solution was applied by him and by S. Kawakatsu in a number of studies, mainly in international comparisons of agricultural production.

Before turning to the details of the Geary–Khamis (hereafter referred to as GK) method, one should note that both Khamis and Kawakatsu were dealing primarily with problems of multilateral comparisons. For them bilateral comparisons are a special case of multilateral comparisons. Nevertheless it seems worthwhile to consider this method in the present context.

The bilateral variant of the GK method is

for price indices:

$$\frac{\sum \frac{q_B q_A}{q_B + q_A} p_B}{\sum \frac{q_B q_A}{q_B + q_A} p_A}$$

for quantity indices, the value ratio divided by the above price index.

Of course, quantity indices can also be compiled directly, on the basis of average prices determined by means of linear equations. However, for the purposes of the present article it is perhaps not necessary to reproduce here this more complicated procedure.

Comparing the GK index with the inter-spatial Fisher formula, the following can be said:

(a) There is no difference in respect of the characteristicity and factor reversal requirements. Both indices satisfy them equally well.
(b) The GK formula satisfies the internal consistency requirement, which is not satisfied by the Fisher formula.
(c) The GK formula is less unbiased than the Fisher formula.

Only this last point needs some explanation. Two features of the GK indices should be noted in this context. The first, admitted by Khamis himself in

[8]R. C. Geary; "A Note on Comparisons of Exchange Rates and Purchasing Power between Countries", *Journal of the Royal Statistical Society*, 1958, Part 1, Vol. 121, pp. 97–99.

his article, is that the result of his index may be outside the Laspeyres–Paasche limits. If one accepts that the Laspeyres index has an upward bias and the Passche index a downward bias, an index which may be outside these limits cannot claim to be unbiased. It is true that one does not know *a priori* in which direction the bias goes and one has also to admit that—as experience has shown—in most cases the GK index remains within the Laspeyres–Paasche borders. However, even in Khamis' own computations there are examples where his index is higher than the higher or lower than the lower of the Laspeyres and Paasche indices. And one should also note that even in cases where the GK index is within the Laspeyres–Paasche limits, it may be much nearer to one of the basic indices than to the other.

The second feature is perhaps even more important. The average prices which are determined by the linear equations are weighted averages where the size of the countries (or more precisely the size of the countries' quantities) also play a role. Thus if a big country is compared with a small country, the average price proportions are nearer to the big country's price proportions than to the small country's price proportions. Consequently, the GK quantity index in this case will be nearer to the index computed on the basis of the big country's prices than to the index using the small country's prices.[9] If an infinitely great country is compared with an infinitely small country, the GK quantity index will be the same as the index using the big country's prices. It would be difficult to deny that there is a bias in such cases. In general, one may say that the GK index has an upward bias for the smaller countries and a downward bias for the bigger countries.

The question whether the size of countries should or should not be taken into account is a more general one and arises not only with the GK formula but with any other formula in which average weights are used (e.g., it can be raised also in connexion with the Edgeworth–Marshall formula). The weighted averages (i.e., when the size of the countries is taken into account) conform more to reality; one may say that these averages do exist in reality (in the ensemble of the two countries), while the unweighted averages are less real. On the other hand, however, one may argue that it is more important to be unbiased than to use more realistic weights.

Coming back to the GK–Fisher confrontation, one may say that, from a theoretical point of view, it is a matter of consideration whether one attributes relatively greater importance to the requirement of unbiasedness or to that of internal consistency. Not only may personal views differ on this point, but—as was already pointed out—the principal aims of a given comparison may also have a determining role in deciding which requirement should be given priority.

## III. The Methods of Multilateral Comparisons

In the preceding section, dealing only with problems of bilateral comparisons, we could completely ignore the requirement of circularity. Now, when we have

[9]This means at the same time that the GK price index will be nearer to the index computed on the basis of the small country's quantity structure than to the index computed with the big country's quantities as weights.

more than two countries, circularity becomes one of the most important factors. In a general way, we shall consider here how circularity can be reconciled with the other requirements dealt with in the previous section.

Before turning to details, it should be stressed once again that the particular circumstances of given comparisons may have a determining influence on the methods to be selected. Situations may vary greatly from case to case, and it is perhaps worthwhile to make a distinction between main types of situations.

One may distinguish first *closed* comparisons from *open* comparisons. In a closed comparison the participating countries are known in advance, and no other country will join it. This may be the case, for instance, for comparisons within a certain community (intra CMEA; intra EEC, etc.). In an open comparison, it is not known in advance how many countries will eventually take part; in principle any country may join it. A special case of open comparisons are bilateral comparisons, which will be multilateralized only at a later stage (e.g., the productivity comparisons carried out under the auspices of the Conference of European Statisticians).

A distinction should also be made between *full scale* comparisons and *limited scale* comparisons. In the former, every country is directly compared with each country participating in the project; in the latter only a limited number of direct comparisons are made, which, however, may cover the ensemble of the countries by means of indirect comparisons (i.e., by chaining the results of the direct comparisons). Whether in a given project a full scale or limited scale comparison is made depends to a great extent on the resources available for carrying it out; however, the extent of interest also plays an important role and the role and the nature of the aggregates to be compared too (it is easier to carry out a full scale comparison in the field of agriculture than in the field of machinery, industry or construction).

Finally, one should distinguish comparisons according to whether the *central interest* is stronger or weaker. There might be very centralized comparisons, where the particular binary interests are to be subordinated completely to the central interest (e.g. intra-community comparisons needed for intra-community planning); in other cases, however, the binary interests may be strong, the binary comparisons are the principal aim of the study, and the comparison for the total is, in a way, only a by-product.

Let us first consider the methodological problems of full scale closed comparisons. We shall examine later what special problems arise when the comparison is not full scale and/or not closed.

*The Closed Full Scale Comparisons*

In general terms, the major issue to be considered is: how can one satisfy the circularity requirement, without doing too much harm to the other requirements?

The main conflict is, of course, that between characteristicity and circularity. As explained earlier, these two requirements are incompatible with one another. One has to make a choice among the following solutions:

(a) To sacrifice circularity completely for the sake of maximal characteristicity, i.e., to accept the fact that results of bilateral comparisons will not

be consistent with each other. This possibility is mentioned only for the sake of completeness; as to my knowledge, it has never occurred in practice that circularity has been completely renounced.

(b) To sacrifice characteristicity completely for the sake of maximal circularity. This solution is more likely to occur, either because those who carry out the study do not care about characteristicity, or because they subordinate it to the central interests. This solution is applied mostly in comparisons conducted or carried out by international organizations. They have a clear central interest (the ranking of countries according to various indicators) and particular binary interests are only faint, if there are any. All comparisons which are made on the basis of international average prices belong to this group. The computations described by Kawakatsu are examples of this.

(c) A compromise: to achieve circularity but without completely sacrificing characteristicity. There are several ways in which such a compromise can be achieved:

(i) *To have two sets of indices*: one which is completely characteristic, the other satisfying the circular test. At first sight this solution seems to be very attractive, since it enables both of these important requirements to be met. However, one has to pay the price that in this case for each binary comparison, there will be two results: a "characteristic" one and a "circular" one. This may cause confusion to readers who are not acquainted with the subtle methodological problems of index numbers, and this is perhaps the reason why this solution does not enjoy great popularity.

(ii) *The Eltetö-Köves or Szulc method* (hereafter referred to as EKS.) The two Hungarians and the Pole, almost at the same time and completely independently from each other, proposed a method by which circularity can be achieved paying the least possible price for it in respect of characteristicity. The least possible price here means that for the ensemble of the comparisons the deviations of the EKS indices from the characteristic binary (Fisher type) indices is minimized (least square method). It is not perhaps worthwhile reproducing the proofs here; the final formula of the EKS indices is, however, surprisingly simple. Each EKS index is a weighted geometric average of the given characteristic index (with double weights) and any possible combination of two chain indirect indices (the ones with single weights). Having country $A$, $B$, $C$, and $D$, denoting the characteristic indices by $B/A$, $C/A$, etc. the EKS indices ($B/A^*$, $C/A^*$ etc.) are determined as follows:

$$B/A^* = \sqrt[4]{B/A^2 \cdot (B/C \cdot C/A) \cdot (B/D \cdot D/A)}$$

$$C/A^* = \sqrt[4]{C/A^2 \cdot (C/B \cdot B/A) \cdot (C/D \cdot D/A)}$$

etc.

In general:

$$I/J^* = \sqrt[n]{\pi\,(I/K) \cdot \pi\,(K/J)}$$

Where $K = 1 \ldots k, j, \ldots n$.

The EKS indices satisfy the circular test: e.g., $B/A^* \cdot C/B^* = C/A^*$. The EKS indices can be applied, of course, not only *instead of* the characteristic indices but also together with the characteristic indices. Thus, the EKS indices can constitute one of the two sets in the "two sets of indices" solution (see c(i) above).

(iii) *The van Yzeren methods.* It is not so easy to determine to which group these methods belong, but perhaps this is the most appropriate place to deal with them. The Dutchman van Yzeren, almost a decade before Eltetö, Köves and Szulc, proposed three sets of inter-country indices, which all satisfy the circularity requirement. There is a striking similarity between the EKS and van Yzeren methods: in all cases the Fisher formula plays the central role. However, van Yzeren sets requirements for his indices other than minimizing deviation from the characteristic indices. His interest focuses on price indices (currency ratios) and he determines his indices with the aim of minimizing the total increase in costs due to ill-adapted currency ratios. All three methods he proposes have their strict economic meaning and may be very useful in special inter-country comparisons. In general purpose comparisons, however, where the main objective is to determine differences in levels between the countries compared, the EKS requirement of minimizing deviation from characteristic indices seems to be more important than the requirements of the van Yzeren methods.

(iv) *The central country solution.* It may happen that in a multilateral comparison the various bilateral comparisons are not considered to be of equal importance. In these cases, while trying to achieve circularity, one may sacrifice more of the characteristicity of the less important bilateral comparisons, than of the characteristicity of the more important bilateral comparisons.

This situation may arise if one single country is the main or only one carrying out a multilateral comparison. In this case one can understand that the given country is more interested in the characteristicity of those bilateral comparisons where it participates itself than in the others, where it is only a "spectator". One could hardly object, if the given country is not willing to sacrifice anything from the characteristicity of its "own" comparisons, while at the same time it is inconsiderate in respect of the characteristicity of "foreign" comparisons. With this consideration circularity can be achieved by keeping all "own" indices fully characteristic, and by compiling "foreign" indices as products or quotients of the appropriate own indices. Having four countries $A$, $B$, $C$, and $D$, $A$ playing the central-country role, denoting by $A/B$, $B/C$ etc. the characteristic

indices, the set of indices which satisfies the circular test ($A/B^*$, $B/C^*$ etc.) is the following:

$$A/B^* = A/B \qquad B/C^* = A/C \div A/B$$
$$A/C^* = A/C \qquad B/D^* = A/D \div A/B$$
$$A/D^* = A/D \qquad C/D^* = A/D \div A/C$$

The central-country solution was applied in some non full-scale comparisons where—owing to lack of all characteristic indices—there was no other possibility of compiling all the indices. This was the case in a number of intra-CMEA comparisons.

Comparing the central-country solution with the EKS method one may say that in both cases the characteristicity sacrifice is distributed among the various bilateral comparisons; however, while in the EKS method this sacrifice is smaller (minimum property of the EKS indices) and its distribution is "neutral", in the central-country method the sacrifice is somewhat greater and its distribution is purposefully determined.

(v) *The regional compromise.* One may often encounter an attitude which is willing to accept a certain extent of uncharacteristicity, but not more. Using again the Netherlands–Belgium example, there might be views which would accept the carrying out of this comparison on the basis of average EEC weights, but not on the basis of averages of the weights of a wider range of countries. What limit is chosen, from a general point of view, is not important. The problem would be the same if Western European weights were accepted but, not all-European weights, or if all-European weights were accepted, but not world weights. In general, if the scope of the participating countries is wider than the acceptable limits of uncharacteristicity, a regional compromise in respect of the characteristicity-circularity conflict may be expedient.

Region is interpreted here as the ensemble of countries which are willing to accept the uncharacteristic influences of each other. We are considering here only those cases where the region is smaller than the total ensemble of participating countries, but comprises more than two countries.[10]

The regional compromise is nothing more than that regional weights (e.g. regional average prices) are used within each region. In this way circularity is easily obtained within the regions, but at the sacrifice of characteristicity. It should be noted that this sacrifice is not so big as it would be if total ensemble average weights were used (e.g. EEC average prices are better for a Netherlands–Belgium comparison than European average prices and definitely better than world average prices). This is certainly a gain as compared with the method using average weights for all the participating countries.

[10]If the region in this interpretation is equal to the ensemble of participating countries, we are back to case (b) above. If the region comprises only two countries, the situation is the same as in case (a).

On the other hand, there is a loss because in respect of inter-regional comparisons, there is no circularity. Or if one establishes in some artificial way this circularity, then intra-regional and inter-regional indices will not be consistent with each other. Thus, the regional compromise, as any other method, cannot solve the characteristicity-circularity conflict either. It only puts it in a different place. Nevertheless, this may also be an important advantage, since it may transfer the difficulties to a place where they are less troublesome. The regional compromise was applied in some comparisons of agricultural production. In the Kawakatsu article one also finds not only world average prices but also European average prices and Asian average prices.

Having concentrated entirely on the characteristicity-circularity conflict, we neglected in this section the behaviour of the different methods from the point of view of the other requirements. Now we have to remedy this. The most expedient way is perhaps to proceed with the requirements.

(a) *Internal Consistency.* Of the remaining requirements, this is the one that deserves most attention. As already explained, the Fisher formula does not satisfy it; thus any solution mentioned which is associated in one way or the other with the Fisher formula, e.g. the EKS indices, the van Yzeren method, the central-country solution, suffers from the internal consistency illness.

This illness is especially serious in the case of indices which are compiled indirectly as products (or quotients) of two other indices. This problem is the same as that of chain indices in inter-temporal measurements. In several studies in which chain indices with moving weights were applied, the overall index was in fact not infrequently outside the range of the sub-indices (i.e. the average test was breached).[11] In international comparisons where the central-country method is applied this may happen in respect of the indices which are computed indirectly (referred to above as "foreign" indices).

Average price methods, like the GK method—or any other method which uses the same weights throughout the whole field—are of course free from internal consistency problems.

(b) *Unbiasedness.* What was said on this requirement in connexion with bilateral studies applies equally to multilateral comparisons. No method which is based on the Fisher formula (EKS, van Yzeren, central-country) is biased; on the other hand the GK formula and all other average weight methods so far known are biased.

(c) *Factor Test.* Nothing important is to be added to what was said on this requirement in the preceding section.

*Special Problems of Limited Scale Comparisons*

In most cases resources do not permit carrying out a full scale comparison (i.e. comparing every country with every other country by direct comparisons). If

[11]As already mentioned, theoretically a non-chain Fisher index is also able to breach the average test; however, in practice the probability that this will happen is much higher in the case of chain indices.

only a limited number of direct comparisons are possible, the question arises as to how they should be selected.

To connect all countries with one another, for $n$ countries, $n - 1$ bilateral comparisons are needed. On the basis of what principles should these comparisons be selected from the total number of bilateral comparisons (which is $n(n - 1)/2$)?

(a) *Coverage completeness.* This means that each country should participate in at least one bilateral comparison, and that it is possible to link any country to any other country. This is a self-explanatory requirement.

(b) *Successivity.* This principle of selection is based on the empirical finding that the smaller the differences in structures between the countries compared, the better results one may expect, i.e., the greater the comparability. Thus, one should pair the countries so that the structural differences within each pair should be as small as possible. To take a fictitious example, if one has Denmark, Federal Republic of Germany, Iceland, Italy, Kenya and Libya to be compared and not more than five bilateral studies are possible, then the best pairing seems to be Iceland–Denmark, Denmark–Federal Republic of Germany, Federal Republic of Germany–Italy, Italy–Libya, Libya–Kenya. Any other pairing would probably be inferior, e.g. an Iceland–Kenya comparison would not be very promising owing to the great structural differences and relatively very low number of common products between the two countries.

(c) *The extent of the many-sidedness of countries.* This principle is similar to the preceding one, but does not necessarily go in the same direction. In general, the more comparable products one has (i.e. products which are produced or consumed in both countries) in a bilateral comparison, the more reliable the indices may be considered. On the basis of this consideration—other conditions being equal—it is expedient to give preference to comparisons with countries which have a many-sided character (which are generally big countries) because they are more likely to produce (or consume) many of the products which the other countries are producing (or consuming).

This consideration often leads to the central-country method or similar solutions. If there is one big country and a number of small countries in a group of countries, the selection of the big country as central-country has the considerable advantage that in this way, the number of comparable products will be certainly higher, since the big country is more likely to produce (or consume) a relatively high proportion of the products which are produced (or consumed) in the other countries. In a number of CMEA comparisons, the U.S.S.R. was the central-country. For similar reasons, in the UN comparison project the United States is playing the role—though not entirely—of central-country[12].

(d) *The interest of countries.* This is, of course, not a methodological criterion; however, its influence on the selection is often far more important than that of the others. If countries have limited resources, they certainly want to give preference to selecting that or those other country(ies) in respect of which they are more interested. One is more interested first, of course, in countries with which one has closer economic ties. A second component of interest is the rela-

[12]The United States is participating in most of the bilateral comparisons; however, there are some other bilateral comparisons too.

tive economic level: one is generally more interested in comparing oneself with countries having a higher level of economic development than with countries which are less developed (one can learn more from the more developed countries, the future development of the own-country will bring some structural changes by which one moves closer to the structure of the more developed countries, etc.).

Of course, if countries were interested only in comparisons with more developed countries in a very strict sense, there would not be a single joint study in the world. Fortunately, one is also interested in making comparisons with countries which are at about the same economic level, or which are significantly less developed, but with which one has close economic ties. Nevertheless divergencies in interests form one of the major obstacles in the development of inter-country comparisons. International organizations which try to promote inter-country comparisons often encounter the phenomenon that there is a willingness in a number of countries to make comparisons, but one cannot find suitable partners for them.

Another problem of limited scale comparisons is connected with the fact that a number of indices are compiled indirectly, i.e., as products or quotients of two or more direct indices. These indirect indices fit in very well from the point of view of the circularity requirement. However, they have some weaknesses: they are not fully characteristic and what is more important, they have serious internal consistency problems (as already mentioned, they can breach the average test in a drastic way). It may happen that in the case of comparisons between countries for which only indirect indices are compiled, both countries have a particular interest in each other and would like to make the indices more useful for their binary purposes, even if the price to be paid could be a loss in circularity. In this connexion, the proposal of E. Krzeczkowska[13] deserves attention. Krzeczkowska suggests that in these cases the linking of two countries should be carried out at a detailed level of aggregation; and that the indices of the more condensed groups and of the totals should be compiled on the basis of the linked detailed indices, using the weights of the two countries concerned only (i.e., without the influence of weights of the intermediary country(ies)).

This solution is, of course, again a compromise. It makes the indices considerably more characteristic (reducing at the same time the internal consistency difficulties); but at the loss, to the same extent, of circularity. If the detailed level of aggregation referred to above were maximally detailed, i.e. if linking were made at the level of individual products, one would arrive at indices which are completely characteristic but not at all circular. In Krzeczkowska's proposal, linking is made at an intermediate level of classification. Now, within each group at this intermediate level, we find properties of circular indices (they are completely circular, but not characteristic, and seriously lacking in internal consistency); and outside these groups (i.e., at a more condensed level of classification) one finds properties of characteristic indices (characteristicity, less difficulties as regards internal consistency, but no circularity). As with many other compromise

[13]E. Krzeczkowska: "The International Comparisons of Consumption Level Carried Out by the Polish Central Statistical Office", *Review of Income and Wealth*, Series 13, No. 4, December 1967, pp. 353–366.

solutions, the Krzeczkowska method might also be very useful in particular situations.

*Special Problems of Open Comparisons*

It may happen that after having carried out the calculations for a group of countries, a new country or several new countries join the project. What impact will this have on the results already obtained? The situation is different depending on the method of multilateral comparisons used.

The smoothest is the situation where the central-country method is used. With this method, the participation of one or several new countries will not change anything in the earlier indices; it will only add some new indices to those already available.

Using the EKS method, the joining of one or more new countries will change all of the EKS indices already available. The characteristic indices on which the EKS indices are based will, of course, still remain valid, but the weighted averages of these characteristic indices (since new characteristic indices have also entered the scope of computation) will be different. The additional work involved as a result of the participation of the new country(ies) is relatively small. The situation is similar with the van Yzeren methods.

With the GK method, the situation is worse. Not only will all earlier indices change, but all earlier computations will need to be done again, even if only one new country is added.

Of course, it would be possible with any method to keep the already-obtained results unchanged and to link them in some way with the new results. But—except for the central-country method—this artificial solution would be a "discrimination" against the newcomers. They would not be treated in the same way as the original countries, and this "stepchild" fate may have serious disadvantages from the point of view of the indices for these countries.