

INTRODUCTION:
THE SPECIAL CONFERENCE ON PURCHASING
POWER PARITIES

BY T. P. HILL*

OECD

At the 17th General Conference of the International Association for Research in Income and Wealth held in Gouvieux in August 1981 it was decided to initiate a series of Special Conferences at which particular subjects could be considered in more depth. It was agreed that the First Special Conference should be devoted to the subject of Purchasing Power Parities and the ensuing Conference was held in Luxembourg from September 22 to 24, 1982. It is expected that, in general, the Special Conferences will take place in the alternate years in which there are no General Conferences.

The Conference consisted of five half-day sessions finishing at midday on Friday September 24, 1982. The number of papers per session was less than in the General Conferences in order to allow more time for general discussion. The number of participants was 35.

The articles published in this issue of *THE REVIEW OF INCOME AND WEALTH* all consist of papers presented at the First Special Conference. Some other papers discussed at the Conference overlapped with papers presented at the preceding General Conference at Gouvieux and have already been published in the *REVIEW* or elsewhere. For this reason, the collection of papers published in this issue represents only a fraction of the papers discussed at the First Special Conference and by no means reflects the full range of papers presented or topics discussed.

THE SELECTION OF ITEMS AND THE CALCULATION OF DETAILED PARITIES

The first two sessions were devoted to the calculation of parities for detailed categories of expenditure or basic headings. Final domestic expenditures are usually broken down into two or three hundred detailed headings as the first stage in the calculation of PPPs. Within each heading it is necessary to specify individual goods and services for which prices are to be collected in order to calculate a parity for that heading. Within a basic heading weights for individual items are not generally available, so that the data do not allow index numbers to be compiled at this level. Nevertheless, the underlying relationships between prices and quantities which give rise to traditional index number problems are still present.

The first two problems on which the Conference focussed were the criteria to be used to select the items to be priced within each basic heading and the

*Chairman of the Organizing Committee.

methods to be used to process the individual price data once they are collected. Particular products which are common or popular in one country may be found only with some difficulty, or even not at all, in some other countries. Moreover, products which are not typical tend also to be relatively expensive so that a negative correlation is usually found between relative quantities and relative prices in different countries, even within a single, narrow category of expenditure. Confining the selection of products to those which are typical of country A may therefore yield a very different parity from one based on products which are typical of country B, or C, etc. One solution to this problem is to ensure that the list of items to be priced within each basic heading includes at least one item which is typical of each country. This procedure has been followed by the Statistical Office of the European Communities, or Eurostat, in its work. While it may be feasible for a small group of relatively homogeneous countries it may pose practical problems for a large group of heterogeneous countries. Unfortunately, experience suggests that the results obtained within a basic heading can be quite sensitive to the selection of items (much more so than inter-temporal comparisons where it is generally fairly easy to observe the price of exactly the *same* item in the same location in successive periods of time). The impact on the final results may be much greater than the subsequent choice of aggregation method.

Participants were acutely aware of the difficulties involved in making an appropriate selection of items, but these are problems which do not admit easy solutions in practice. The more similar countries are to each other, the easier it is to match individual items, which suggests that more consideration needs to be given to constructing chains in which the links always involve comparisons between fairly similar, but not necessarily neighbouring, countries. While there is no automatic procedure by which countries can be put into an ordered sequence (again in contrast to inter-temporal comparisons) there may be ways of devising acceptable criteria for this purpose.

When the list of items to be priced has finally been decided, there still remains the problem of processing the raw price data into a set of *transitive* parities. This problem arises mainly because not all items can be found in all countries so that the lists of prices actually supplied by different countries, although overlapping, may vary significantly from country to country. It also arises if it is decided to ignore certain prices as being irrelevant, or inappropriate, for comparisons between particular pairs of countries. Two methods of achieving transitivity are in current use, namely the so-called CPD and EKS methods.¹ They tend to give similar results, as the paper by Mr Krinjse-Locker shows, although this need not always be the case. Transitivity is achieved in both cases by allowing the parity between any individual pair of countries to be marginally influenced by prices in other countries.

A further problem which can arise at the level of the basic heading occurs when it is difficult to lay down precise specifications for individual items because

¹For a description of these methods, see the Reports on Phases I, II and III of the United Nations *International Comparisons Project* (ICP); the results of Phase III are given in Irving B. Kravis, Alan Heston and Robert Summers: *World Product and Income* (Johns Hopkins University Press, 1982). See also T. P. Hill: *Multilateral Measures of Purchasing Power and Real GDP* (Eurostat, 1982). See also *Comparison in Real Values of the Aggregates of ESA, 1980* (Eurostat, 1983).

their physical units are not easily defined. This applies to many services, especially government services in the fields of health, education, public administration and defence. When these services are provided free there are no price observations anyway, so that the comparisons must start from quantities instead of prices. There was a lengthy discussion of these problems which have recently come to be regarded as particularly important for international comparisons of real GDP based on PPPs. Although the problems of "comparison resistant" services are by no means peculiar to international comparisons, their consequences are felt to be more serious than in inter-temporal comparisons. The conventional solutions adopted in inter-temporal comparisons of basing the output measures on inputs with an assumption of equal productivity yield results which many users find unacceptable in an international context, especially between countries at very different levels of economic development. For example, it may be disputed whether a typical doctor or teacher in a poor country can be nearly as "productive" as a typical doctor or teacher in a rich country who may not only be better trained but also be provided with better equipment and facilities. It is argued that methods taken over from inter-temporal measurement tend to over-estimate the production and consumption of many non-market services in poor countries relatively to rich ones. While there may be some truth in this, in the absence of proper quantity measures it is not clear that the biases involved are so great as is often alleged. Moreover, the measures actually in use are being steadily improved and refined either by the development of genuine output measures (e.g. treatments or pupil-hours of tuition actually provided), or by adjustments to take account of estimated differences in productivity. The Conference was informed about improved output measures for non-market services which are being developed within the European Community.

AGGREGATION METHODS

The aggregation methods used in international comparisons provoked the most controversy and the liveliest discussion at the Conference. Although the international organisations mainly involved in this work, the United Nations Statistical Office, Eurostat and the OECD, have agreed to use the so-called Geary-Khamis method² it is not universally accepted that this is the best method available at the present time. For this reason, a more detailed description of the issues involved will be given in the following paragraphs.

There is much greater demand for volume than price measures at an international level. While there may be some groups of people who have a direct interest in international price comparisons as such—people who work abroad or travel a lot such as businessman, diplomats and international civil servants, and tourists—PPPs are mainly intended to be used as international price deflators for National Accounts data expressed in national currencies. The resulting volume measures can then be used to make meaningful comparisons of living standards or productivity.

²See Irving B. Kravis, Alan Heston and Robert Summers, *op. cit.*: T. P. Hill, *op. cit.*: Eurostat, *op. cit.*

In practice, the procedures traditionally followed in National Accounts have been followed. Volume measures are constructed by measuring goods and services in different countries at constant prices, i.e. a common set of international prices, the detailed PPPs calculated for the basic headings being used to revalue flows expressed in national currencies. Users are already familiar with constant price data from time series of national accounts. Such data are not only easy to understand and to manipulate, but are also amenable to econometric analysis. Data for different countries can also easily be aggregated to obtain National Accounts for groups of countries, such as the EEC or the OECD. Moreover, the use of a common set of international prices automatically ensures that the volume measures are transitive at every level of aggregation. The main debate, therefore, has centred not on whether or not to use constant price data, but rather on which set of prices to choose for this purpose.

Before considering this question, it is worth noting that the consequences of using one set of international prices rather than another are well understood from index number theory and practice. The more closely the pattern of international relative prices resembles the actual pattern of relative prices found within a given country, the lower the volume measure for that country will tend to be *relatively* to other countries. This follows from the well known tendency for the relative prices and relative quantities for any pair of countries to be negatively correlated. The index number spread between Laspeyres and Paasche indices is a special case of this tendency: when the international prices are identical with those of the base country in a binary comparison the volume measure for the base country will tend to be lower, relatively to the second country, than when the prices of the second country are the international prices. It can, therefore, be predicted in advance what are the implications of using one set of international prices rather than another. For example, if average European prices are used to compile a set of OECD measures, the share of Europe in the total GDP of the OECD area will tend to be lower than if North American (or Japanese) prices are used. In itself, however, this kind of knowledge does not help to determine which are the most appropriate prices to choose.

One school of thought argues that the natural choice of international prices for a set of *multilateral* measures is simply the average prices within the group of countries *as a whole*. These are the prices which most users would consider most relevant for their purposes. By definition, they are likely to be more similar to the relative prices at which most transactions take place within the group than, say, the average prices of some arbitrarily selected individual country or group of countries. The average price of a commodity is given by the total value of the transactions in that commodity divided by the total quantities transacted (i.e. a simple, unweighted mean of the prices at which all the individual units of that commodity are sold) but, to obtain such an average over a group of countries, values expressed in different currencies have to be converted into a common numeraire. This could be done in various ways, but following an original idea of R. C. Geary,³ the preferred method in practice is to use the PPP for a broad aggregate, such as GDP, as the currency converter.

³Geary, R. C., A Note on Comparisons of Exchange Rates and Purchasing Power between Countries, *Journal of the Royal Statistical Society*, 1958, pp. 97-99.

More precisely, if PPP_j is the purchasing power parity from currency j into the chosen numeraire, the average price π_i for commodity i is:

$$\pi_i = \frac{1}{\sum_j \sum_k q_{ijk}} \cdot \sum_j \left(\sum_k \frac{(p_{ijk} q_{ijk})}{PPP_j} \right)$$

where commodity i is sold at k different prices within country j and q_{ijk} is the quantity sold at price p_{ijk} .

A commodity does not sell at uniform price within an individual country and, when defining the average international price, it is essential to recognize that the price of a commodity may vary significantly between different locations, or sales outlets, within the same country. Indeed, it is precisely this local variation in prices, combined with the fact that there is usually no way of matching individual sales outlets in one country with those in other countries, which creates one of the main practical problems in the calculation of PPPs. On the other hand, local variation in prices is easily handled in inter-temporal comparisons because repeated observations can be made on prices in the same outlets in different time periods which enable price ratios, or relatives, to be calculated directly at the level of the individual sales outlet. However, there is usually no way of directly linking individual sales outlets in different countries in the same period of time, so that national average prices have to be compared in international comparisons. The calculation of these national average prices can sometimes pose serious data problems, especially in large countries.

There is a second school of thought, however, which relegates the international prices to a subordinate, or instrumental, role and seeks to evaluate multilateral measures in terms of their relationships to binary measures. Two main propositions are advanced. First, it is argued that a binary measure between two countries taken in isolation from the rest of the group is *ipso facto* the best measure for that pair of countries and therefore the yardstick against which to appraise any multilateral measure. When binary comparisons are made systematically within a group of countries additional measures are implied between any given pair of countries by the binary measures with third countries: e.g. direct measures between A and C and between B and C yield an implicit, or indirect, measure between A and B, and so on for countries D, E, F, etc. According to the second school of thought, any modification to the original direct measure to take account of the considerable additional information contained in these implicit or indirect measures is regarded as automatically detracting from the quality of the direct measure. The greater the divergence between the multilateral measure and the original binary measure, the worse the former is regarded as being.

In general in statistics, however, the utilization of additional information is not regarded as reducing the quality of estimates. For example, the utilization of the information contained in the indirect measures may well actually improve the direct measures at a world level, especially between pairs of countries which are far apart economically as well as geographically where there may not be sufficient overlap between the items available in both countries. In such cases, the quality of the binary comparisons may not only be improved but significantly improved by taking account of comparisons with intermediate countries which

can serve as links to bridge the gap between the two original countries, especially when this is done systematically and objectively within the framework of a set of multilateral measures.

The second proposition is that countries have to be treated symmetrically, a proposition which appears eminently reasonable at first sight. The basic data consist of a pair of vectors for each country, a vector of quantities and a vector of prices, and the elements of these vectors are treated as if they were direct observations on quantities and prices in different locations to which equal weight or importance should be accorded as a matter of principle. It follows from this second proposition that the best binary measure is an index such as Fisher's which treats both pairs of vectors symmetrically. This, together with the first proposition, elevates Fisher's index to the status of *the* best measure for any pair of countries in some absolute sense. It is sometimes asserted that other measures are "under" or "overestimates" to the extent that they deviate from the Fisher index or even that they are "biased" estimates.⁴ It must be noted, however, that these terms are used in quite a different sense from that which is now normally understood in modern statistical theory and for that reason they can be quite misleading. In any case, for reasons outlined below, advocates of the first school of thought certainly do not accept that the Fisher index, in the context of *inter-country* comparisons, is the best measure in any absolute sense.

If, however, there are circumstances in which the Fisher index is the best binary measure, the best set of transitive multilateral measures becomes one which minimizes the sum of their deviations from the corresponding binary Fisher indices, and this is the rationale for the so-called "EKS" method.⁵

The EKS method has one serious disadvantage in this context, however, namely that it is not additively consistent. Other measures can be devised, however, which yield almost identical results to the EKS method and which are additively consistent. One method is to define international prices which consist of simple, unweighted geometric means of the national prices: another is to choose as the vector of international prices one which is equi-distant from the vectors of national prices. Both methods have been proposed by D. Gerardi.⁶ The first method is easy to understand and, in the case of a binary comparison, has certain attractions over the Fisher index itself because it is conceptually simpler and more meaningful, while yielding almost identical results in practice. It will, therefore, be used here as the prototype of the second class of measures under consideration.

Modern index number theory is usually formulated in terms of vectors of prices and quantities relating to *points* in time or space.⁷ In practice, temporal index numbers relate to short periods of time of equal duration which can be regarded as discrete approximations to moments of time. The spatial equivalent of temporal indices is to compare vectors of prices and quantities for different locations which, in practice, have to relate to small areas of equal size, such as circles of specified radius around the centres of capital cities—the areas within,

⁴See, for example, Eurostat, *op. cit.*, p. 47.

⁵See Irving B. Kravis, Alan Heston and Robert Summers, *op. cit.*: T. P. Hill, *op. cit.*: Eurostat, *op. cit.*

⁶See Eurostat, *op. cit.*, pp. 41, 42.

⁷See, for example, Afriat, S. N., *The Price Index* (Cambridge University Press, 1977).

say, a radius of two or three kilometres. This example is not hypothetical as various organizations seek to make comparisons of prices (or living standards) between the centres of large cities for the determination of appropriate salary levels and allowances for their employees. In this situation, it is quite legitimate to treat the vectors of prices and quantities observed in the different locations symmetrically as the locations are all defined in the same way and are therefore all commensurable with each other. Thus, the EKS method, or the first Gerardi method, may be perfectly suitable for the purpose of calculating a set of multi-lateral price of cost of living indices relating to specific locations in different countries.

However, countries are not discrete approximations to points in space. Nor are they units of fixed dimensions. On the contrary, countries cover geographical areas which vary enormously in size and whose boundaries are quite arbitrary from an economic point of view. Moreover, because the boundaries are arbitrary they can readily be changed, if desired, by splitting large countries into smaller units for purposes of internal comparisons or by grouping smaller countries together for purposes of comparisons with large countries. For example, one may wish to compare the GDP of New York State with that of California or to compare that of the European Economic Community with the United States. Thus, when the observations relate to flexible areas which can be manipulated in this way, it is not even clear how many sets of observations are involved, precisely because it is meaningless to ask how many areas there are within a given country or group of countries.

The vectors of prices and quantities for the United States, for example, patently do not consist of single observations. They consist, in fact, of a vector of *aggregate* quantities and a vector of *average* prices relating to expenditures over an enormous area. (It has already been emphasized in connection with definition of the average international prices used in the Geary–Khamis method that there is considerable variation in prices within countries and that the inputs into international comparisons are not national prices but national *average* prices.) Thus, the vectors of quantities and average prices for the United States can meaningfully be decomposed into at least fifty separate pairs of vectors of quantities and average prices for the individual States, many of which are far larger, both economically and geographically, than many Nation States. Such a decomposition might be needed for a variety of political or economic reasons. Conversely, the vectors for the individual member countries of the EEC can be consolidated into a single pair of vectors of aggregate quantities and average prices. This kind of decomposition, or consolidation, is frequently undertaken for statistics relating to areas of variable dimensions. On the other hand, there is little point in trying to decompose or consolidate observations which in principle relate to quite distinct points in time or space, especially when they are not even adjacent to each other. It is highly questionable, therefore, whether it is appropriate to use a theoretical model which treats the vectors of quantities and average prices for different countries as if these were original observations, relating to different points in time or space all of which have to be treated symmetrically.

It is quite meaningful to compare the volumes of the GDP's originating within areas of vastly different size, or to aggregate these GDP's, or to disaggregate

them, since GDP is a concept which is specifically defined with respect to a geographical *area*, whether large or small. On the other hand, an average price becomes progressively less useful as the area covered is enlarged because the purpose of price comparisons is to compare levels of prices in different locations, and the local variations in prices are gradually averaged out as the area is enlarged. A comparison between the average prices in Europe and North America, for example, is a great deal less interesting or useful than a precise comparison, say, between the centre of Paris, or London, or Rome and New York. In general, international price comparisons become progressively less useful, in themselves, the larger the areas to which they relate, whether those areas are located in one or several countries, even though the associated PPPs may remain valid as National Accounts deflators.⁸

Consider, for example, a comparison between the GDP of the United States and that of the EEC, a comparison of genuine economic interest. Eleven countries are involved, of which ten are in Western Europe and one in North America. If the international prices are defined in such a way that the average price for each country, defined as a separate *political* unit, carries an equal weight, it follows that the average prices for the EEC as a whole will implicitly be carrying about ten times as much weight as the average prices for United States. It is difficult to justify such procedure when, in practice, the aggregate GDP of the EEC is about only four-fifths that of the United States. One might wish deliberately to choose a procedure whereby the pattern of international prices is more or less the same as the pattern of average prices within the EEC, but not on the grounds that such a procedure is neutral, objective and “unbiased”. It is clear that such a procedure will tend to underestimate the GDP of the EEC as a whole relatively to the United States, as compared to a method which assigns equal weight to average EEC and average United States prices.

The root of the problem is that it is difficult to attach any meaning to the notion of equal weighting, or symmetry, when dealing with arbitrary, modifiable geographical units. Is the United States one observation or an aggregate of fifty or more observations? Is the United Kingdom one observation or four observations? There is, for example, at least as good a case from an economic point of view for treating Scotland (which also happens to have its own bank notes as well as its own language, and traditions and economic structure) as a separate entity as, say, Luxembourg. Conversely, within an OECD framework, should the EEC, or the Nordic countries, be treated as blocs for purposes of comparison with large non-European countries, in particular, the United States or Japan? The distribution of GDP within the OECD area as a whole ought not to be sensitive to whether or not the GDPs of the member countries of these blocs are aggregated first prior to comparison with other countries.

Thus, advocates of the first school of thought argue that the international prices should, so far as possible, be invariant to the way in which the boundaries of the countries which make up the group are drawn, or redrawn. It should not

⁸Lazlo Drechsler and Eugenia Krezczkowska have also emphasized the difference between PPPs as price measures and PPPs as national accounts deflators, but on other grounds: see *Purchasing Power Parities in International Price Comparisons: Quantity vs. Price Comparisons*, *Review of Income and Wealth*, series 28, No. 3, September 1982.

be possible to increase the weight given to the prices in one particular country, or area, by a factor of n simply by subdividing it into n smaller components. Nor should the weights attached to well defined sub-groups, such as the EEC countries within the OECD group of countries, be greatly affected according to whether they are treated as composite group or as separate countries. Marked variations in weights for reasons of this kind are quite arbitrary and inconsistent with objective, scientific measurements. The best way to avoid the introduction of arbitrary weighting schemes seems to be to define the international prices in such a way that economic activities and transactions are accorded equal weight wherever they occur within the group, in the same way that average prices are defined within countries.

CONCLUSIONS

Disagreements between the two schools of thought provoked lively discussion without either side convincing the other. However, it may have led to a better appreciation of each other's point of view. For example, supporters of Geary-Khamis pointed out that the EKS method, or the Gerardi method, may well be quite suitable for a system of multilateral prices indices referring to the centres of different cities, or other precisely defined locations, but argue that the problem in hand differs in a number of important respects from generalized index number problems of this kind. The issue is not simply an index number problem but also one of compiling national accounts data across countries of widely different sizes in an objective manner which makes economic sense and which will be generally accepted by users. It is also clear that, despite the enormous progress made since the start of the ICP project, a number of methodological problems remain unresolved, especially at the level of the basic headings, and that much work remains to be done.