Volume Title: Output, Input, and Productivity Measurement

Volume Author/Editor: The Conference on Research in Income and Wealth

Volume Publisher: Princeton University Press

Volume ISBN: 0-870-14181-3

Volume URL: http://www.nber.org/books/unkn61-1

Publication Date: 1961

Chapter Title: International Comparisons of Real Product and Productivity
by Final Expenditures and by Industry

Chapter Author: Milton Gilbert, Wilfred Beckerman

Chapter URL: http://www.nber.org/chapters/c2225

Chapter pages in book: (p. 251 - 274)

International Comparisons of Real Product and Productivity by Final Expenditures and by Industry

MILTON GILBERT AND WILFRED BECKERMAN
ORGANIZATION FOR EUROPEAN ECONOMIC COOPERATION

THE GROSS national product per capita (or per unit of labor input, if that is more suitable to the purpose), adjusted for price changes, provides a convenient over-all measure of intertemporal variations in the productivity of an economy. We in the OEEC have worked along these lines to obtain comparable indexes for our various member countries and for the OEEC area as a whole, so that the rates of change among the countries may be compared. The essential problem, of course, is to construct measures of real gross national product and of the labor force for the various countries that are conceptually and statistically similar. Our annual reports and publications on national product and expenditure show that considerable progress has been made in this work, although improvement of the data for some of the countries is still needed.

In this paper, however, we are concerned with another type of work designed to yield, not international comparisons of rates of change, but international comparisons of the absolute level of real GNP per capita or per unit of labor input. The essential idea is to provide the same type of basic data as that yielded by intertemporal comparisons, but with the indexes relating to a given point of time and the index numbers showing the comparative standing of various national economies.

We have approached this work by two methods. In the first, the basic datum for the construction of the indexes is the GNP broken down by final expenditures. Two reports have been published using this method. One in 1954 covered five countries and gave data for 1950 and 1952; one in 1958 covered nine countries and gave data for 1950 and 1955.[1]

In the second method, the basic datum is the GNP broken down by industry of origin or net output. This study, by Deborah Paige and

---

[1] Milton Gilbert and Irving Kravis, *An International Comparison of National Products and the Purchasing Power of Currencies* (OEEC,. Paris, 1954); and Milton Gilbert and Associates, *Comparative National Products and Price Levels* (OEEC, Paris, 1958).

Gottfried Bombach, was published in June 1959 as a joint project of the OEEC and the Department of Economics of Cambridge University under the title of *A Comparison of National Output and Productivity of the United Kingdom and the United States*. In addition, two papers dealing with this method have been issued.[2]

Needless to say, the work with both these approaches was undertaken in an experimental spirit and intended to show the kinds of problems encountered, as well as the character of results that could be obtained. In this paper we intend to focus more on the problems that have not been dealt with adequately in previous papers or reports.

The Conceptual Basis of the Comparisons

THE CONCEPT OF PRODUCTION

The conceptual basis of the expenditure comparisons has been discussed in some detail in the first of the OEEC publications mentioned above. The following discussion will be confined mainly to those aspects of the conceptual problems that have been questioned or which have been the subject of further work since the publication of that study. It was said that the primary aim of the study was "to compare the flow of goods and services made available for the purposes of consumption, gross investment, and government by the country's own production resources."[3] It was emphasized that such comparisons are subject to well-known limitations. Most of these limitations, which are of various kinds, apply also to measures of the changes over time in a country's real national product. They are such that comparisons of real national product must not be confused with comparisons of well-being or happiness in a wider philosophical sense.

Nevertheless, measures of relative real national product are of considerable significance. Other things being equal, the size of the real national product is a component of well-being in a wider sense, and is the component about which economists and politicians are able to do something. Other elements of happiness, particularly those of a spiritual nature, are no doubt very important, but economists are not supposed to be responsible for them and, in any case, they are not subject to measurement. Whatever reservations may be attached to the concept, people do happen to be closely interested in the size of the

[2] Deborah Paige and S. Adler, "International Comparisons of National Products: An Approach by Industry of Origin," *Income and Wealth*, Series IV (1955), and Milton Gilbert, Gottfried Bombach, and Deborah Paige, "The Real Product of the United Kingdom and the United States in 1950," paper presented to the Fourth Conference of the International Association for Research in Income and Wealth at Hindsgavl (1955).

[3] Gilbert and Kravis, *op. cit.*, p. 61.

national product, its growth, and its composition. As economists our duty includes both the satisfaction of this interest and the provision of the statistical measures of this concept, which are indispensable tools of economic analysis. However, there are divergent views regarding the appropriate concept of output to use in productivity comparisons and, in fact, a question in some minds as to whether international comparisons can have a firm foundation at all.

For purposes of "real" comparisons of total national output, there are three concepts to consider:

1. The consolidated output of domestic establishments. This is the gross domestic product which includes exports but excludes imports.

2. The output of domestic establishments plus net income from abroad. This is the gross national product, including exports but excluding imports.

3. The gross national product made up of private consumption, government consumption, domestic capital formation—all including imports but excluding exports—and net foreign investment (including net income to or from abroad). This concept is identical to concept 2 at current prices, but for comparisons over time or between countries at constant prices they are not necessarily so. Concept 3 corresponds, in intertemporal comparisons, to GNP adjusted for changes in the terms of trade. It has been suggested that this concept be called the gross national income.

We have used this third concept in the OEEC studies and believe strongly that it must be considered the basic concept of aggregate production for both intertemporal and international measures. The reason for this view is that it measures the final output of the national economy in consumption and investment.

Technical and analytical objections have been raised against this viewpoint. The technical argument in favor of the first concept is that any index of production must be made up only of goods (i.e., commodities and services), the quantities and prices of which are the basic data for the calculation of the index. No purely financial flows, such as net income from abroad or net foreign investment, should enter into the index as these do not have specific flows of goods behind them. Such flows can only be deflated by attributing a flow of goods to them which would introduce an arbitrary element into the whole index. Thus the only logical concept of aggregate real output is the gross domestic product.

While this argument has formal validity, we do not believe it to be decisive. It may be more straightforward to determine how much the gross domestic product differs from situation A to situation B, but

253

this is not the question we want answered. We want to know how the national product, adjusted for any changes in the terms of trade, differs in the two situations because only that tells us what the residents of the nation are getting out of the economic process. If, in fact, the domestic product differs significantly from the national product, then it yields a result that cannot be interpreted, either from a productivity or from a welfare standpoint, as one cannot say either who produced the output or who received it.

Hence, we must try to find an acceptable technique for deflating net foreign income or investment, as the case might be. This can readily be done by using an import price index for net income received from abroad and for net foreign borrowings, or by using an export price index if the signs are reversed. This method is certainly appropriate since imports and exports are the real flows that correspond to the financial flows. The defect that remains is that one cannot specify which imports or which exports and is forced to use a general import or export price index. But this defect does not seem great in the light of the many compromise techniques that must be used to deflate aggregate output altogether. Moreover, one can say that the larger the financial flow is proportionately and the more important it is to deflate it to get a meaningful result, the more will the general indexes of export or import prices approximate the real flow of goods that should be measured.[4]

The analytical argument in favor of the gross domestic product is that the idea of productivity itself is only relevant to a specific flow of goods and services from a specific group of establishments. Presumably, productivity in this context, being conceived of as a kind of measure of industrial efficiency, would only be distorted by the addition of net income from abroad or by differences in the terms of trade.

This point of view seems to us to be wide of the mark essentially because a measure of overall productivity (either intertemporal or international) is not simply a measure of the difference in industrial efficiency narrowly defined. One of the main elements in its variation is differences in the industrial composition of output in the two situa-

[4] In some discussions of this problem, the impression is given that the choice of the real flow to represent the financial flow depends upon the purpose of the measurement. For example, it is said that net foreign investment might be deflated by import prices because the economy gives up the opportunity of obtaining additional imports by the act of investing abroad. Or, that net foreign investment might be deflated by an index of consumption goods prices because the investors forego additional consumption by the act of investing. However, when the real gross national product as a whole is being calculated, such considerations are irrelevant; each component of the gross national product can only relate to a particular category of real output. Net foreign investment can equally be called net exports and, hence, the only appropriate deflator to make the total result meaningful is export prices.

254

tions being compared. Foreign investment and foreign trade are essential parts of the economic process and a comparison of two situations leaving out these elements is incomplete.

So far as net income to or from abroad is concerned, we see no reason for omitting it from the calculation and consider that results without it are generally not very useful. Suppose, for example, that an economy has substantially increased its output with the help of a large import of foreign capital resulting partly in a flow of income to abroad. To measure the increase in productivity of this economy without allowing for this import of capital would be like measuring the increased productivity in agriculture without allowing for the increased use of fertilizers and machinery. While this may have some meaning in isolation, it has no meaning as part of a measure of overall productivity.

With regard to the effect of differences in the terms of trade, this has considerable independent significance, particularly in intertemporal comparisons of a single economy. This significance largely revolves around the analysis of year to year changes. However, the idea of productivity tends to lose much of its relevance in periods of cyclical change. From a longer-term standpoint when productivity is compared at the same stage of the business cycle, say at full employment, then the difference in the terms of trade cannot be considered an aberration. For example, the British economy produces less of its food requirements because it finds it advantageous to get these from abroad in exchange for manufactured goods. To say what the productivity change of the British economy has been after this shift in its structure, it is surely essential to include foreign trade as part of the productive process. Similarly, if the United States and British economies are being compared, only the final products available to the two economies will give a meaningful result.

DIFFERENCES IN TASTES AND INCOME DISTRIBUTION

Following the first of the OEEC studies mentioned above, J. L. Nicholson raised certain considerations relevant to the question of how far the comparisons provide valid indicators of relative economic welfare or productivity. He stated that "the problems involved in comparing the national products of two different countries are essentially different from those involved in comparing the national products of a single country at two points of time."[5] He notes that the main considerations are differences in tastes and in income-distribution between countries.

[5] J. L. Nicholson, "Official Papers: The International Comparison of National Products," *Economic Journal*, June 1955.

Nicholson also mentions that, between countries, "capacities for enjoyment" of a given collection of goods may differ and that "the enjoyment of economic goods depends to a large extent on uneconomic factors—the proximity of friends and relatives, the life-long associations of a particular part of the world, etc."[5]

It is necessary to specify what is meant by differences in tastes in comparing two countries. For purposes of the welfare theory of index-number comparisons, it is essential that the community pre-ference fields, described by indifference surfaces relating quantities of goods purchased to their relative prices and to varying levels of in-come, be very similar. In this context, differences in tastes must be interpreted as the action of variables other than prices and incomes which result in different collections of goods being purchased at a given level of relative prices and incomes.

Thus we are not concerned with differences in, say, the style or design of the goods purchased. If one country has a greater preference for coffee than for tea—quite apart from differences in incomes and relative prices—such a taste difference is only significant if the two products require different types and quantities of real resources. In such a case, whether tea and coffee are treated as separate products or whether they are combined (after making appropriate quality adjust-ments), the quantities consumed would show differences between the countries beyond those that may be due to differences in incomes and relative prices. In other words, the indifference curves of the two countries cross whether or not tea and coffee are treated as separate products and, strictly speaking, no valid welfare comparisons are possible.

From the point of view of productivity comparisons, however, even this sort of taste difference may, in practice, be of minor importance. This would be the case when the *relative* real resources required to produce the same competing goods were the same in both countries. For example, suppose that, as a result of "taste" difference, rela-tively more radios were purchased in the U.S. than in the U.K., but U.K. purchases of sewing machines were relatively greater than in the United States. Then suppose that in both countries a given num-ber of sewing machines could be produced by the resources required to produce some other given number of radio sets. (This is assuming there would be no change in relative costs if the proportion of the two goods produced were changed.) The relative positions of the exchange lines in the two countries tells us nothing about the relative positions of their indifference curves but it tells us something about the relative positions of their production transformation curves. Given the above

[5] Nicholson, *op. cit.*, p. 255.

assumption of a constant rate of production substitution between sewing machines and radios, these are parallel straight lines when drawn in two dimensions only (i.e., when relating only possible output of sewing machines to possible output of radio sets). For example, at the same level of aggregate consumption of radios and sewing machines, the exchange lines would coincide, even though the relative proportions of the two goods purchased in the two countries would differ.

From the welfare point of view, two questions arise: (a) whether there are likely to be greater differences in tastes between countries than between two years in a given country and (b) whether in any particular comparison, the differences are likely to be so great as to invalidate conclusions that might otherwise be reached. It does not seem to the present writers that any definite general answer can be given to the first question, nor does it seem that, with regard to tastes, there is any essential difference of principle between international comparisons and intertemporal comparisons—it is purely a matter of how far apart the two situations being compared are. Comparisons between two periods of time which are far apart (in terms of the scope for changes in tastes) may be just as limited as comparisons between two countries. It could be argued, however, that comparisons over very long periods of time are invalid as measures of changes in welfare, but that this does not make the international comparisons any less invalid, since these too will break down on the obstacle of large differences in tastes.

But what is the evidence for the assumption that, in general, international differences in tastes are greater than intertemporal comparisons involving, say, similar differences in income levels? In the second of the OEEC studies, an experimental analysis was made of the variations in consumption patterns among the nine countries covered in the study, with reference to differences in their incomes and their relative price structures.[6] In spite of the fact that in 1950 the play of consumer preferences was still far from free in some of the countries, or was subject to special distortions arising from postwar recovery, it was remarkable how greatly the observed international differences in consumption patterns appeared to be explained by differences in incomes and relative prices. These results suggest that international differences in tastes are not so important as is sometimes alleged, on the basis of no emipircal evidence of which we are aware. This result is not particularly strange. If the average Frenchman buys more wine than the average Englishman, it is more likely because wine is relatively cheaper in France than in England, rather

[6] Gilbert and Associates, *op. cit.*, Chapter V.

than because every Frenchman is born with a taste for it or that every Englishman is born with a preference for beer. It is true that a low price of wine would result in a habit of wine consumption and that a subsequent rise in the price might take a long time to have its full effect. But this is merely a defect in the statistical methods of demand analysis employed which the use of more dynamic demand functions would eliminate. Furthermore, such failures of demand to adapt itself to sharp changes in relative prices or incomes affect only an insignificant minority of the products concerned.

There may be, in fact, reason to believe that international differences in taste are *less* important than differences over long periods of time in one country. Countries are, after all, connected by international trade and by many vehicles of communication so that the average Frenchman knows perfectly well that the average American has a big car, a television set, and drinks whiskey. It was much more difficult for the American of 1900 to know what the American of 1950 would consume, or to consume it if he had wished to. It is not out of ignorance that the average Frenchman drinks relatively little whiskey, and it is not for climatic reasons that the average Italian has no refrigerator. It is largely because their incomes are lower and because whiskey is relatively expensive in France and refrigerators are relatively expensive in Italy. Thus, there seems to be no *prima facie* reason for believing that international differences in tastes are particularly marked; such empirical evidence as is available suggests that they are not.

There remains the question of whether small differences in tastes invalidate the comparisons as indicators of relative economic welfare. According to the present state of the theory of welfare economics, the greater the difference between the two situations with respect to their real national products (or levels of consumption) the greater the certainty, other things being equal, that the apparent statistical difference reflects a real welfare difference. As I. M. D. Little has put it: "If one (person) wishes to maintain that consumption has risen and the other (person) that consumption has fallen, neither can possibly *prove* his case. Nor could he prove it however perfect the statistical techniques might be. Of course, if the index number showed a terrific rise, the person who maintained that consumption had fallen would look silly. One might cease to trust his judgment, but he could never be proved wrong. If, however, the index number showed only a small rise, it really becomes quite meaningless to say that consumption has either risen or fallen."[7]

It has been argued above that there is no reason in principle to

[7] I. M. D. Little, *A Critique of Welfare Economics* (Oxford, 1950), p. 218.

believe that there are greater differences in tastes between countries than there are over long periods of time within a given country and that, in fact, there are reasons to believe the contrary. Furthermore, the magnitude of the apparent differences in real levels of GNP or consumption as measured in the studies concerned are very large in some cases. Even the U.K. appears to have had (in 1955) a *per capita* consumption level about 50 per cent below that of the U.S. If value judgments have to be made—as they inevitably must in this type of comparison—it would be a bold man who could maintain that such apparent differences in *per capita* GNP's are likely to be significantly invalidated by differences in tastes between the average Englishman and the average American.

Similar considerations apply to the question of income distribution. It is true that to make welfare comparisons, some value judgment has to be made concerning income distribution. In the OEEC studies, abstraction has been made from the question of income distribution. The comparison of the relative products of the countries indicates the relative welfares that could be achieved if income distribution were no better or worse in some countries than in others. Given the magnitude of the differences found between the national products of the countries concerned, however, it would again be a bold man who would assert, as a value judgment, that they were significantly affected, from the point of view of welfare, by offsetting differences in the goodness of the income distributions.

QUALITY DIFFERENCES AND UNIQUE PRODUCTS

The main conceptual obstacle to the validity of the comparisons thus seems to be the question of unique products (products appearing in one situation but not in the other) and of quality differences. As Mr. Little has stated, "the introduction of a new good formally prevents us from saying that welfare has increased, and, similarly, if a good ceases to be produced, it would prevent us from saying that welfare has decreased. Quality changes also can obviously invalidate the analysis."[8]

In practice, of course, techniques can usually be found for assimilating the unique products into other commodity classes by means which do not appear to do too much violence to the rationale of the calculation or to the actual statistical results. Nevertheless, there is a limit, and when unique products become too numerous there may not be much of substance left in the comparison. On this point we have said, that "the necessity of having prices or costs of unique goods for both markets as data for the index-number equations point to the

[8] Little, *op.cit.*, p. 218.

fundamental difficulty involved in comparing the national products of very primitive and advanced economies. This difficulty has often been seen primarily in terms of the difference in tastes and needs. From the point of view adopted in this report, however, the difficulty is rather the large amount of unique goods in the national expenditure of the advanced society for which no meaningful price or cost can be imputed to the primitive society. The primitive society has no possibility of producing the good itself; an estimate of relative costs would be fictitious. Moreover, it has no possibility of importing the goods in significant quantities without a complete upheaval in its domestic price structure, assuming that this possibility exists at all. Hence the index number equations can be computed only on the basis of the price weights of the advanced society. This result may tell us something either from the standpoint of productivity or economic welfare."[9]

Quality differences give rise to a similar conceptual problem. They must be treated in practice partly by abstracting from them and partly by adjusting for them. There seems to be general agreement on how adjustments for quality differences should be made, so we may here discuss only what quality differences must be neglected in a quantitative measure of a change in production.

First, it is essential to draw a distinction between what we have called economic and non-economic differences in quality. Statistical adjustments can be made for the former whereas none can be made for the latter since they are essentially non-quantitative in character. We can define a non-economic change in quality as one which comes about through the general improvement of science and the arts but which does not involve any increase in production costs. For example, a physician of today can render a better quality of service to his patients than one of twenty-five years ago, simply because medical science has advanced (leaving aside drugs, medical equipment, and hospital services, which are products separate from the physician's services). There is no way to bring this improvement of quality of medical services into an index of production, just as there is no economic aspect of it that can serve as a basis of measurement. Hence, as in an international comparison, indexes of production must abstract from such non-economic quality changes.

Second, the whole question of what constitutes an improvement in quality involves an element of judgment and, therefore, the process of constructing an index of real production consists partly of an appraisal of what is or is not significant. If one takes the view that every difference between a given product in the two situations is a

[9] Gilbert and Kravis, *op.cit.*, p. 87.

quality difference, then the whole task becomes hopeless. And this is the case even if the differences exist in the base year (or country) with differences in prices attaching to them. For example, consider the case of an identical product which sells for a different price in chain stores and in independent stores, or the case of a product of identical physical specifications which sells at a different price packaged and unpackaged. We believe that as a matter of judgment, these products must be considered to be without difference in quality and that the price differences represent imperfections of the market. Of course, it can be argued that the consumer gets something extra from the convenience of packaged goods or from the convenience of buying in a neighborhood store. But it seems more reasonable to consider such differences as changes in the process of production and distribution which are undertaken as a matter of business efficiency and not as affecting the consumer's economic welfare. Without a simplifying assumption of this kind, it really becomes physically impossible to construct any index of overall production.[10]

When one abstracts from the two kinds of quality differences mentioned above the task of dealing with the quality problem is statistically manageable. Nonetheless, the resulting index of production is subject to additional conceptual limitations from the standpoint of both welfare and productivity. We do not feel that it is necessary to throw out the baby with the bath water, but there are evidently extreme situations in which there is not much of a baby left.

PROBLEM OF WEIGHTS

It has been argued above that, given a substantial difference in the real national products of the two situations compared, valid conclusions may be drawn since other conceptual limitations (tastes, income distribution) will, in most cases, be relatively unimportant. Unfortunately, the greater the disparities in real national products, the greater the differences between the price (and quantity) structures of the two situations tend to be, and hence the greater the effect of the choice of weights on the measure of the disparity in real products (or in purchasing power equivalents). For example, as can be seen in Table 1, the relative position of the United Kingdom's *per capita* GNP on U.S. price weights is 29 per cent above its relative position on its own price weights, whereas for Italy it is 67 per cent above. How

---

[10] The whole matter of what constitutes a change in quality is not very clearly treated in the literature. For example, in the generally excellent book of Erland von Hofsten, *Price Indexes and Quality Changes*, Stockholm, 1952, there does not appear to be a precise definition of what a quality change really is. Evidently he believes it is of very frequent occurrence from the perspective of constructing a cost of living index. We believe a much more restrictive view is necessary for production indexes.

TABLE 1

Indexes of per Capita Gross National Product of Eight European Countries in 1950
(United States per Capita GNP = 100)[a]

Country	At U.S. Price Weights (1)	At National European Price Weights (2)	Ratio of (1) to (2) (3)
United Kingdom	62	48	1.29
Denmark	61	48	1.27
Norway	59	44	1.34
Belgium	57	48	1.19
France	53	39	1.36
Netherlands	52	37	1.41
Germany	44	30	1.47
Italy	30	18	1.67

[a] Gilbert and Associates, *op. cit.*, Table 4, pp. 28 and 163.

serious this weighting spread is depends, of course, on the particular
question posed. Clearly, if one is interested in giving precise quanti-
tative expression to the disparities in real product (or the subsidiary
aggregates) the magnitude of the weighting spread constitutes a
serious handicap, and the use of geometric means, for example, how-
ever unsatisfactory these may be from the conceptual point of view,
is no doubt essential. However, this handicap accurately reflects the
underlying economic meaning of comparisons between very different
consumption patterns or price structures, and any method, such as
that proposed by Nicholson, for example, for overcoming this handi-
cap merely involves disguising, rather than solving, what is after all
an insoluble problem.[11]

In the OEEC studies, the important difficulty arises from the
desire to attach ordinal, if not cardinal, relationships between the
pairs of countries compared in the binary comparisons (comprising
separate comparisons of each of the eight European countries with
the United States, both on U.S. weights and on the weights of each
individual European country concerned) as well as among the various
European countries themselves. The problem was whether to derive
such relationships among the European countries from the set of
eight binary comparisons. The objections to this procedure have been
discussed by Gilbert and Associates,[12] who point out that this involves

[11] Nicholson proposes to base the comparison on an initial selection of goods which
have the same relative prices and the same quantities of consumption in both situations.
Apart from the impossibility of knowing what "relative prices" are until the over-all
price level has been established, this method involves pretending that the two situations
are nearly the same in order to avoid the problems arising out of the fact that they are
different.

[12] Gilbert and Associates, *op. cit.*, pp. 153–5.

262

comparing pairs of countries on the basis of a set of weights which is common to neither. For this reason an alternative procedure was adopted which involves "Average European Weights," representing a weighted average of the price structures of all the European countries covered in the study.

This procedure is still subject to important limitations. "It is perfectly possible, for example, that Average European Weights, which are relatively heavily influenced by the weighting structures of the larger countries, including Italy, will have even less in common with the weighting structure of, say, Norway and Denmark, than would United States weights, or perhaps Australian weights (or even some imaginary set of weights). But while this may be the case for one or two European countries, it is not likely to be so for the eight European countries taken as a whole, not only because they are all represented to some extent in the computation of the Average European Weights, but also because, as is mentioned above, the United States income level lies far outside the range of income levels in the eight European countries."[13]

In fact, to compare any pair of European countries, the only solution is to make separate binary comparisons between them using the weights of both. This technique has been adopted in a recent study by the European Coal and Steel Community in which a complete matrix of binary comparisons between the six member countries of the Community has been established.[14] As the authors of this study would acknowledge, even this procedure does not provide any unique overall ranking of all the countries concerned, as the impossibility of doing so in a manner which provides a satisfactory set of common weights for the whole ranking is inherent in the nature of the index-number problem.

As far as questions of conceptual or theoretical principle are concerned, we conclude that the conceptual limitations on intercountry comparisons of national product are no greater, and may be even less, than those which apply to intertemporal comparisons. In both cases limitations exist on account of differences in tastes, income distribution, quality differences, etc. But it does not seem to the present writers that these considerations seriously affect the order of magnitude of the results that may be obtained. This is particularly so on account of the fact that even if there were no differences in tastes, some degree of uncertainty must be attached to these results on

[13] Gilbert and Associates, *op. cit.*, pp. 154 and 155.

[14] See publications by the European Coal and Steel Community, Luxembourg, *Informations Statistiques*, July–August 1957, and *Comparaison des Revenus Réels des Travailleurs des Industries de la Communauté: Analyse Statistique*, 1957.

account of the weighting problem, not to mention purely statistical problems.

The Expenditure Approach and the Industry of Origin Approach

ARE THERE TWO INDEPENDENT METHODS?

Mention has been made in the introduction to this paper of the comparison between the real products in the U.K. and the U.S. on the basis of the net outputs of the various sectors contributing to GNP. The preceding studies compared the prices and volumes of the goods and services entering into final demand. The industry of origin approach involves comparing, sector by sector, the value added by the sector concerned, first at the prices ruling in one country and then at the prices in the other. In principle, the comparison at, say, country A's prices, is made by estimating value added in the various sectors in country B at country A's prices, which involves, again in principle, deducting from the total output of each sector in country B valued at country A's prices for the products concerned, its total inputs of materials, etc., also valued at country A's prices. In practice, the complete quantity and price data necessary to apply this technique were not available and various simplifications had to be adopted. First, though, it is useful to consider the extent to which the expenditure approach and the industry of origin approach are independent.

If one conceives complete input-output tables for two situations being compared, and if all the transactions shown for situation 1 are valued at the prices prevailing in situation 2, the sum of the value added, shown in the first table, will still be equal to the sum of the final demands in that table—both being at the prices of situation 2. This is simply a mathematical identity, because all intermediate products cancel out, whatever prices one uses. In both cases all that is left is the value of the transactions entering into final demand, even though, by the industry of origin method the aggregate value of these transactions appears as distributed among the sectors in the form of their net outputs.[15] The industry of origin approach thus produces the same answer that would be obtained by the expenditure (i.e., final demand) approach since it is this answer which remains after all intermediate transactions are cancelled out. In this sense, therefore, the industry of origin approach is in no way independent of the expenditure approach, since if the answer by the latter were not already put into the calculation to start with, the cancelling out of

[15] For a formal exposition of this result see Richard Stone, *Quantity and Price Indexes in National Accounts* (OEEC, Paris, 1956), pp. 39–44.

all intermediate transactions would not necessarily leave behind an aggregate equal to GNP.

In order to value the net output of sector A in country A at country B's prices, it is necessary to include in the valuation of sector A's output its sales to final demand at country B's prices. To make such a valuation for all sectors, one has to include the sales by all sectors to final demand, each valued at country B's prices. But these constitute all that is necessary in the expenditure approach.

This point is stressed here because its implications are considerable in practice since one usually does not have all the data necessary to carry out a complete deflation of all outputs and inputs in the manner described above. This applies to both international and intertemporal comparisons.[16] In that case, the problem is to know what significance can be attached to comparisons based on net outputs when gaps in the data have to be overcome, particularly by use of the assumption of constant input coefficients. It would appear that the industry of origin approach gives a "correct" answer only insofar as it gives the same answer as the expenditure approach.

Now it is frequent in practice (e.g., in indexes of industrial production) to overcome the absence of complete data on inputs and outputs by assuming constant input coefficients. For if input coefficients were to remain constant, then net output at constant prices would move parallel to gross output or to gross inputs (whichever indicators are used). But insofar as this assumption is not realized (and, in general, it is not) any comparison of GNP's by the industry of origin approach based on this assumption will give an answer which is "wrong" in the sense that it is not the same as that which would be given by an expenditure comparison, if one could be made.

This raises the question of the advantages of the industry of origin approach. Either the data enable a complete deflation to be carried out, in which case one has to have all the data necessary for the expenditure approach, or one does not have the full data, in which case the practical expedients adopted to overcome the deficiencies will tend to introduce errors. The answer lies, in fact, largely in the sphere of the relative statistical problems encountered in the two approaches. One particular analytical advantage of the industry of origin approach arises from the fact that it is usually desirable to obtain not merely a comparison of aggregate GNP, but also of its components.

16 The results of intertemporal comparisons made in the U.K. have appeared in various issues of the *Bulletin* of the London and Cambridge Economic Service (published quarterly in *The Times Review of Industry*). The methodological problems are discussed in W. B. Reddaway, "Movements in the Real Product of the United Kingdom, 1946–49," *J.R.S.S.*, Series A, 1950, Part IV, and "Some Problems in the Measurement of Changes in the Real Geographical Product," in *Income and Wealth*, Series I (1951).

In the expenditure approach, for example, the analytical value of the comparison is not limited to the results concerning relative aggregate GNP's. It is particularly interesting to compare also, in real terms, the structures of the GNP's both in respect to quantities and to prices. For example, one important aspect of the two OEEC studies was the light the expenditure comparisons threw on the relative "real" contribution made by the different countries to the common defense effort.

From this point of view, it is obvious that even though the result for aggregate GNP given by the industry of origin approach is logically dependent on expenditure data, the additional information on the contribution of individual sectors and the derived information on relative productivities may be of considerable value.[17]

Unfortunately, this advantage may be exaggerated. The industry of origin approach is only correct if full allowance is made for every sector for the differences in input coefficients. But this will produce results for the individual sectors which, while giving a correct total for GNP, may have limited economic significance in themselves. In the extreme case, the valuation of outputs and inputs in one situation by the prices of another situation can give negative value added. And if constant input coefficients are assumed, then while the results for the individual sectors appear more definite, the GNP will be distorted.

This feature of the industry of origin approach, however, is not entirely absent from comparison of individual items in the expenditure approach. In the former, the limitations on the results for individual industries arise not only because one is dealing with balancing items, valued at the prices of different situations, but also because there may be several inputs into one sector and the pattern of inputs varies with their relative prices. In the expenditure approach, insofar as a given item is not perfectly homogeneous and includes several subitems, the comparison between the aggregates for the item valued in the prices existing in only one of two situations is also subject to limitations of the type discussed in the preceding section of this paper. In other words, wherever there is a lack of homogeneity, there is no simple answer to the question of what are the relative aggregate amounts used in two situations. The use of index numbers provides only a limited answer, whether it is a problem of adding together consumers' inputs of eggs and cheese or of adding together the steel industry's inputs of pig-iron and coal.

The flexibility of the input structure in some industries may be very

[17] For an excellent study of production by this method, see J. B. Heath, "British-Canadian Industrial Productivity," *The Economic Journal*, December 1957.

limited on technological grounds. Smooth iso-product curves are abstractions and in many important industries the range of production techniques available may be restricted to very few combinations of inputs.

RELATIVE MERITS OF THE TWO APPROACHES FOR STATISTICAL REASONS

The relative conceptual merits of the two approaches may be considered both from the standpoint of the extent to which each gives a "correct" measurement of relative aggregate GNP's and from the standpoint of the analytical value of either on the breakdown of expenditures or the contribution to total value added by individual sectors. The former type of information is more interesting for demand analysis; the latter for the analysis of productivity. The reliability of the results realized in practice, however, depends on the statistical problems encountered in each approach.

For the industry of origin approach, the basic information required is the GNP in a breakdown by industry, with appropriate quantity indicators of output and inputs for each industry. The expenditure approach, on the other hand, requires the GNP in a breakdown by final expenditure with appropriate quantity indicators for each category of expenditure. Both approaches draw upon the same basic data for all sectors and types of output, except for manufactured goods. Hence, from the standpoint of reliability, the choice of method for any pair of countries must depend largely on whether the industry or the final product data for manufactured goods are better.

So far we have studied only two countries by both methods—the United Kingdom and the United States. The results are shown in Table 2.

TABLE 2

Gross National Products of the United Kingdom and the United States, 1950, by the Expenditure and Industry Methods
(Quantity Index: U.S. = 100)

| | Per Capita GNP | |
	U.K. Price Weights	U.S. Price Weights
Industry method	55	65
Expenditure method	48	62

The United Kingdom comes out better in the industry study, its *per capita* output being 15 per cent higher (relative to the United States) than in the expenditure study on United Kingdom weights and 5 per cent higher on United States weights. While we consider these differences to be reasonably small, given the nature of the two

calculations and the fact that they are both first attempts, they are not negligible—particularly on the U.K. price weights.

We must confess to having no firm view on which of these two sets of results is statistically better. The industry study was done later, and thus some improvements of both data and method could be brought to it, but this does not do much to reduce the differences. The best we can do is to list elements of weakness and strength of the two methods, given the basic data that were available for both calculations.

There was available for both countries an excellent census of manufactures. A lot of work was required to make the industry classifications comparable and this reshuffling is a possible source of error. Nonetheless, the mass of detailed information on manufacturing gave a firm base to the industry study. (Anyone who has done this kind of work will know the feeling of confidence one gets from having a good census as a point of departure).

For the expenditure approach, on the other hand, a much less detailed breakdown was available. The major components of final expenditures were available in value form, of course, in the usual GNP data. But for many of these categories the detailed quantity and price data required for the comparisons had to be obtained by various inquiries and estimation procedures.

From this point of view, the industry approach seemed to offer two advantages. First, a much greater weight in the comparisons is borne by relatively good quantity indicators for fairly homogeneous semi-finished products—steel, cloth, paper, rubber, etc. Secondly, for certain important categories of final manufactured products, quantity indicators could be obtained from census material for a large part of the aggregates, whereas in the expenditure approach virtually no direct quantity indicators were available. This applied particularly to producers' durables, household goods, and clothing. Consequently, the quantity indicators had to be derived in aggregate from the original value data by means of price comparisons. This involved both the difficult task of collecting representative price data for as many components as possible of the categories and also estimating the appropriate weights with which to combine the price data. In the industry approach, of course, the latter part of the work was relatively unimportant in view of the detailed quantity data available for many of the components. However, the absence of the readily available detailed price and quantity data in the expenditure approach is not a difficulty of principle, and can be overcome according to the time and resources at one's disposal.

Against these advantages must be set its disadvantages. The chief

disadvantage of the industry approach is that in principle a correct result can only be obtained by a complete deflation of all the inputs and outputs of the various sectors. In practice, however, the census of production data which have been used do not include sufficient information on inputs for such a complete deflation, and compromise solutions have had to be adopted. For certain non-manufacturing sectors, particularly agriculture, fuel, and transport, it has been possible to treat inputs and outputs separately (i.e., to use the "double indicator" method), but this has not been possible to any significant extent for manufacturing, the sector on which the relative statistical merits of the two approaches really depends. Only two adjustments have been made for differences between the two countries in input coefficients. One was applied to manufactures in aggregate; this is the adjustment for differences in the relative fuel inputs. It was not possible, however, to make this adjustment for the individual component sectors of manufacturing. The other adjustment made was for differences in transport input; here the correction could only be applied globally to all transportable goods (including manufacturing, agriculture, and fuels).

Thus the inability to allow fully for differences in input coefficients in manufacturing will give rise to error, and unfortunately it is not possible, *prima facie*, to be sure which way the error will lie.

The other disadvantage of the industry approach is its treatment of quality differences. In the expenditure method, an attempt was made to obtain price data for goods of equivalent quality or, where these did not exist to adjust for the difference in quality through an estimate of cost. There is no reason, in principle, why this method should not be fairly satisfactory, in spite of various practical difficulties. The limiting factor is again largely the time and resources available for the investigation. With the industry approach, as carried out so far, it might appear that one aspect of quality differences may be adequately accounted for, namely those arising from differences in inputs of materials (since these will be reflected in the outputs of the sectors producing the materials). But it is not known how far greater inputs of a certain raw material are reflected fully in higher quality rather than in a greater degree of wastage. This problem probably does not apply to many sectors (assuming that, as between countries even more than over time, knowledge of production techniques is fairly similar), but it will no doubt apply to some.

In addition, quality differences that are not due to differences in inputs but to the amount of value added cannot be allowed for in the industry approach without making the same kind of quality evaluations as in the expenditure approach, and, in principle, for a much

269

greater variety of transactions. As, in fact, a greater part of manu-facturing output in the industry method could be covered by direct quantity indicators, there was necessarily less adequate consideration given to quality differences than in the expenditure method.

We hope that more experience will give greater insight into the relative reliability of the two approaches.

COMMENT

TIBOR SCITOVSKY, University of California

There is a tug of war between the more imaginative statisticians and the more rigorous—or should I say more pedantic—theorists. The former have occasionally discovered interesting correlations, estab-lished significant rates of growth, and found suggestive international differences. All too often, however, the rigorists have attacked their methods, questioned their results, and even denied the meaning of the problems they posed. Such criticism is useful if it leads to a better understanding of the underlying concepts and the development of better statistical methods. If it is purely destructive and defeatist, however, then it is largely self-defeating, for it is hard to find a statistical series that has been discontinued owing to its methodologi-cal condemnation by the theorists. Indeed, the field of economics is littered with such live corpses: concepts, laws, and statistics that mathematical economists have condemned to death and the rest of us continue to use because we find them useful.

International comparisons of standards of living are among the very few examples of statistical work that have been abandoned under the methodological onslaught of the theoretical economist. Gilbert's work is the first in this field in almost two decades. Although I have been cast for the role of the pedant for which I am temperamentally ill-suited, I have the greatest admiration for Gilbert's excellent work and for his temerity in doing it at a time when the rigorists have the field.

I agree with Gilbert's and Beckerman's choice of the gross national income concept for international comparisons. I also agree with their analysis of the difference between the final expenditure and the in-dustry of origin approach. I should like, however, to discuss further the meaning of international product comparisons.

As an example, take the finding of Gilbert and his various as-sociates[1] that in 1950 the U.S. *per capita* GNP was 61 per cent above

---

[1] I am referring not only to the paper here under discussion but also to the two OEEC volumes: M. Gilbert and I. B. Kravis, *An International Comparison of National Products and the Purchasing Power of Currencies* (OEEC, Paris, 1954) and M. Gilbert and Asso-ciates, *Comparative National Products and Price Levels* (OEEC, Paris, 1958).

the British figure, taking U.S. prices as weights, and 108 per cent above the British figure, taking British prices as weights.

This can be given either a welfare or a productivity interpretation; I propose to deal with both. As for welfare, the two indexes are estimates of the distance between two levels of consumer well-being, or rather between two indifference surfaces. It is unfortunate that this can be measured at different points, in different directions, and with different results. The 61 per cent can be shown to underestimate the distance between John Bull's and John Doe's indifference surfaces when travelling from the point and in the direction of the British product mix. The 108 per cent is an overestimate, unfortunately not of the same distance along the same route, but of the distance between the two indifference surfaces in the direction and to the point of the U.S. product mix. That the two indexes are estimates of distances along different routes is unavoidable. The best one can hope for is that the two distances are not too dissimilar; and economists usually regard as the best guarantee of this that the two indifference surfaces belong to the same system of integrated wants. Unfortunately, international comparisons offer the greatest scope for the nonfulfillment of this condition.

I suspect, however, that such pessimism is partly a matter of one's frame of mind. Economists, traditionally cast for the role of Voltaire's Pangloss, go to the opposite extreme when they deal with index-number problems; and they usually assume the worst imaginable conditions for comparisons of real income. Perhaps the first to strike a more optimistic note are J. R. Hicks[2] and, to some extent, Gilbert and Beckerman in the paper under discussion. I should like to rationalize the Gilbert-Beckerman approach.

If it were not for differences in the product mix in different countries and at different dates, I think economists would not worry about differences in different people's preference functions but would instead proclaim the basic similarity of human wants. International differences in the product mix can be explained not only by differences in preference functions but also by differences in production possibilities and in standards of living. Indeed, the more the difference in product mix can be explained by these other factors, the more confidently can one assume the similarity of preference scales.

This seems to be the approach followed by Gilbert, Beckerman, and their associates. They use partial regression analysis to explain international differences in product mix by differences in income levels and relative prices, presumably imputing the price differences to

[2] J. R. Hicks, "The Measurement of Real Income," *Oxford Economic Papers*, N.S., Vol. X, June 1958, pp. 125–62.

differences in production possibilities.[3] Having thus explained 85–90 per cent of the differences in product mix, they feel justified— and I think rightly so—in assuming that what they call the average consumer's long-run indifference map is not very different in the different countries. In other words, they assume that the two indifference surfaces belong more or less to the same integrated system of preference scales. All this means, strictly speaking, is that the indifference surfaces will not intersect, although it is customary to read into it the meaning that the distances between the two indifference surfaces will not be too dissimilar along the two routes.

Now, the productivity interpretation. Here we seem, especially at first, to be on firmer ground. We avoid altogether the slippery concept of consumer's welfare and concern ourselves with production and productivity. On this interpretation we want to know the distance between the British and the American production frontiers. The Paasche and Laspeyres quantity indexes now become estimates of this distance measured at different points and in different directions. This looks like the same awkward situation that we faced on the welfare interpretation, but it is not because now the difference between the distances measured along the two routes becomes not only not awkward but also highly significant. The superiority of the U.S. economy over the British is bound to be very different in different fields and industries and these differences are expressed by the shapes of the two countries' production frontiers. The difference between the distances along the two routes is one indication of the nature and extent of these differences. The 61 per cent difference between the British and the American *per capita* GNP now becomes an *over-*estimate of the distance between the two countries' production frontiers in the direction of the British product mix; the 108 per cent figure now becomes an *under*estimate of the distance in the direction of the U.S. product mix. In other words, U.S. *per capita* output would be less than 61 per cent greater than the British if Americans insisted on consuming it in the form the British do. It is more than 108 per cent greater than the British only because Americans have adjusted their expenditure pattern so as to consume more of the goods the American economy is especially good at producing and thus make better use of the latter's special advantages.

I think this interpretation is more meaningful, since it makes use of the discrepancy between the two estimates, a piece of information that is left unused by the welfare interpretation. But unfortunately it, too, has its shortcomings. In the absence of perfect competition, the the production frontiers to which the price planes are tangential

[3] Gilbert and Associates, *op. cit.*, Chap. V.

certainly cannot be given the interpretation of technologically deter-mined production possibility surfaces. They can, however, be given the meaning of Graaff-type production feasibility surfaces, which express the limitations imposed by technological and institutional conditions.[4] The trouble with these production feasibility surfaces is that one cannot assume that they are convex. Indeed, these short-comings of the productivity interpretation explain the preference of most economists in the past for the welfare interpretation.

I should like to point out two things in favor of the productivity interpretation. First of all, the shortcomings of this approach seemed formidable in the past because we felt overconfident about the wel-fare approach. Now that we have fully realized the limitations and shortcomings of the welfare approach, the productivity approach looks, by comparison, more promising. Secondly, I feel much sympathy with Hicks's position[5] that for making real income comparisons we need not know the shape of production frontiers or indifference sur-faces throughout their entire extent but only within the range bordered by the two product mixes; and this range is likely to be fairly narrow.

PAUL STUDENSKI, Special Legislative Committee on the Revision and Simplification of the Constitution

Gilbert has effectively demonstrated that the difficulties in making international comparisons of national income are not necessarily greater than the difficulties in making intertemporal comparisons of it. I wish, however, that he had made one point more clear, namely that the international comparisons in question can be made precisely only between countries that are similarly advanced economically. Such comparisons are misleading when made between economically advanced and economically underdeveloped countries whose pro-duction and consumption are wholly different. Gilbert suggests this, but he does not state it in so many words.

The Gilbert-Kravis international binary comparisons of national income published by the Organization for European Economic Co-operation in 1954 were limited to some five economically advanced countries. I recently asked Gilbert whether in his new work (which I have not seen) he still limited his binary comparisons to economi-cally advanced countries, or had extended them to the undeveloped countries. Gilbert answered that his binary comparisons were still limited to the economically advanced countries; that at one time he had hoped to include Turkey and started work upon it, but had given up because of the lack of sufficient price data and of sufficient staff.

[4] Cf. J. de V. Graaff, "On Optimum Tariff Structures," *Review of Economic Studies*, Vol. XVII, pp. 47–59, where they are called efficiency loci. [5] Cf. *ibid.*