OUTPUT AND PRICES IN THE INTERNATIONAL COMPARISON PROJECT*

BY IRVING B. KRAVIS

University of Pennsylvania,

AND

ZOLTAN KENESSEY

United Nations

In this paper we discuss a few of the problems that have been encountered in defining output and in comparing prices for the International Comparison Project (ICP). We report also on the way in which these problems are being met.

The ICP has for its purpose the establishment of a systematic set of procedures for making international comparisons of gross domestic product (GDP) and of the purchasing power of currencies. Substantive work on comparisons involving Colombia, the European Economic Community (EEC), Hungary, India, Japan, Kenya, the United Kingdom and the United States is also being carried on with the aid of the statistical services of the countries and of the EEC. It is hoped to expand the comparisons beyond these countries as rapidly as possible.

THE MEANING OF OUTPUT

The SNA as the Basic Source of Concepts

The conceptual problems involved in defining the gross domestic product (GDP) so that it constitutes an unduplicated aggregate that distinguishes the results of economic activity from non-economic activity have been discussed now for the better part of a half century. The outcome has been for some questions clear resolutions based on underlying theoretical considerations, while for others, where theory could not resolve the issues, conventions commanding international agreement have been developed. These resolutions and conventions have been carefully and systematically set out in *A System of National Accounts* (SNA).[1]

In the concept of GDP that has emerged, there are, it is well understood, many aspects of welfare that are not measured. Gross domestic product tells us nothing about such important aspects of welfare as the pleasantness or unpleasantness of working conditions, the net improvement or deterioration of the environment, or the equities or inequities of the economic and social organization.

*The ICP is being carried on by the UN Statistical Office in cooperation with a group at the University of Pennsylvania which has been financed mainly by the Ford Foundation. The ICP has also been supported by the International Bank for Reconstruction and Development, and assistance has been given by other organizations including the Economic Commission for Europe, the Economic Commission for Africa and the Brookings Institution. For further information about the ICP, see :"UN International Comparison Project, Methods for International Product and Purchasing Power Comparisons", September 1969 (multilith) hereafter referred to as "Methods . . ." The authors gratefully ackowledge the comments of Alan Heston and Robert Summers in connection with this paper. They have also benefited greatly from discussions on these issues with the statisticians of the countries and organizations cooperating in the International Comparison Project.

[1]United Nations, *Studies in Methods*, Series F, No. 2, Rev. 3 (1968).

What the GDP does do is to measure in a reasonably satisfactory way the changes from time to time in the provision of satisfaction-yielding commodities and services, holding constant the production boundary marking off economic activity from other human activities.

In the context of international comparisons, the GDP concept enables us to judge the relative flow of outputs of satisfaction-yielding products from the economies of two or more countries in an analogous manner to that in which the relative flows are compared for two or more time periods within a country.

The SNA is thus an obvious reference framework for an international comparison of real products. It provides a standard set of definitions and classifications to which all cooperating countries can be asked to conform with respect to the way in which the detailed data necessary to carry out the comparisons are provided. That is, it gives guidance on what should be included and how. What is also sometimes helpful, the production boundary defined by the SNA serves as a reference source to determine in questionable cases what it is that should *not* be included in our comparisons.

Yet a number of conceptual and practical problems remain when it comes to international comparisons. There is first the question of whether international comparisons are justified at all. Secondly the SNA, detailed and careful as it is, necessarily leaves some leeway in its practical application, and the possibility arises that two countries may conform to the SNA and still be left with some incomparabilities in their GDP's. International differences in GDP coverage that may be relatively unimportant for measuring time-to-time changes within countries may have to be taken into account in place-to-place comparisons. Third, there are a number of problems of implementation of SNA concepts and procedures that have not been fully resolved in time-to-time measurements that are also encountered in place-to-place measurement.

The Validity of Making International Comparisons

The utility of and interest in international comparisons of real product are evident in the fact that they have been attempted by many major international organizations (including the UN, OEEC, ECLA, IBRD, COMECON, and ECE), governments (including the U.S.S.R. and the U.S.), and individual scholars. They are variously desired to judge comparative economic performance, for burden-sharing, and for analytical purposes in setting growth targets and in studying growth processes.

The question has sometimes been raised, however, whether comparisons are justified in principle. What meaning can be attached, to put in an extreme form a question that is sometimes asked, to numbers that purport to compare bundles of commodities and services consumed in the villages of Asia and Africa with the radically different bundles consumed in the cities of Europe and America?

This challenge to international comparisons can only be based (unless it is accompanied by an equal challenge to intertemporal comparisons) on the size of the gap between the quantities and prices in the two situations. There is nothing in the logic of the objection that does not apply to some degree to comparisons between two closely adjacent periods for a given country. Comparisons are justified, in rigorous theory, only from the standpoint of a given

person at a given moment in time; one can legitimately ask questions about differences in the money income that would be required to leave a given individual indifferent between the price structure he actually faces at that moment in time and some alternative price structure or structures with which he might be confronted at the same moment. Even comparisons of the welfare of the same individual at two points in time cannot be made rigorously, because it cannot be assumed that his tastes remain identical as he passes through life.[2]

These niceties are, however, generally ignored in intertemporal comparisons. Indeed, when the comparison is confined to a short interval of time such as one year to the next or even a period of several years, few would regard it as unreasonable to treat tastes as unchanged. It is true that the hemline may rise or fall and that car fenders may change their shape, but if one is prepared to accept the desire for changes in style as an inherent part of the utility function once a certain level of income is reached, then such changes need not be regarded as changes in taste.

Over a longer period of time, however, the style of life appears to change in a more fundamental sense. The clothing we wear is apt to be made of different materials, the entertainment we enjoy may reach us in different forms, we may travel from our homes to our places of work and to our friends' residences in different forms of transportation, and even the food we purchase is different in many cases from that available in an earlier age.

It places a still greater strain on our credulity when we make a similar comparison between distant situations in space at a given moment in time. Are we really warranted in assuming that the taste structures of such different peoples as the French, the Japanese, the Indians, and the Americans are similar?

There are two ways that this question—and the related question involving long term comparisons within a country—may be answered. On one level, stress may be placed on the differences in the physical forms of the things people consume in the different situations. French wine, Japanese saki, German beer, and American Coca Cola can be regarded as reflecting different tastes. So, too, can the relatively large quantities of automobiles, refrigerators, and other durable goods consumed by Americans as compared to those consumed by people in other countries. Indeed, products can be found in some countries that have no single equivalent or no equivalent at all in the consumption pattern of other countries. For example, the Kashmiris use a tiny stove filled with glowing coals known as a kangri, next to their persons to keep off the winter chill. It is difficult to think of an equivalent in other countries; some combination of a blanket, pajamas, and sweater would be required to serve the same purpose.

On the other hand, it is possible to regard the basic needs and desires of man as fundamentally the same in different periods and different places. In this view, what changes from time to time and what is different from place to place is not so much what men would like to have but what it is that the economy affords them. What differs is first the extent to which the economy is capable of satisfying their needs, and secondly the means by which they are satisfied—that is, the physical identity of the goods. Differences in technology have been mainly responsible for

[2]Cf. F. M. Fisher and K. Shell, "Taste and Quality Change in the Pure Theory of the True Cost-of-Living Index," J. N. Wolfe, ed., *Value, Capital and Growth* (1968).

variations in the physical forms in which the economy produces goods that satisfy basic wants that have remained substantially constant through time and over space.

There are reasons for believing that it is the second of these views that in the main is the more valid approach to the international differences that are observed in the real income of nations. Support for this view can be found both in ordinary observation and to a modest degree in econometric analyses. Any traveler around the world cannot help but be struck by the similarity of goods found in the shops of the major cities. Indeed, one could be in a department store in Tokyo or in New York and not be able to tell which was which, were it not for the differences in language and appearance of the people. Plumbing facilities in housing and the possession of automobiles and other durable goods, once regarded as American idiosyncracies, are becoming common in other parts of the world as economic levels rise. Even in parts of Asia where material values are often thought to take a second place to spiritual ones, emerging middle classes typically pursue the same patterns of consumption that people with equivalent income levels pursue elsewhere in the rest of the world. The implication is that the consumption patterns of the peasant and subsistence sectors in less developed countries are different from those of middle-income urban dwellers in the same countries and from the consumption patterns prevalent in the richer countries, not because of differences in taste, but because of differences in opportunity. If this is the case, the measurement of relative income levels between the rural and urban sectors of the less developed countries and between the less developed countries as a whole and the rich countries is in principle a valid exercise. The problem then becomes one of finding criteria of equivalence between the different physical forms of goods that are used to satisfy similar wants. This is difficult, but not objectionable in principle and, as we argue subsequently, manageable in practice.

In the realm of econometric and statistical materials, there is no extensive evidence on either side on this issue, but the few hints that are available support the general view that the tastes of men are more alike than not. One clue is given by the tendency for price ratios for subcomponents of consumption and of GDP in country-to-country comparisons to be inversely correlated with the quantity ratios.[3] This is consistent with the existence of international similarities in price and income elasticities. Also, several studies suggest that at least for the United States and the countries of Western Europe, there are great similarities in consumption patterns; income elasticities of demand for various categories of goods and ownership patterns for consumers' durables seem to be alike.[4]

What is Output?

The relationships between what we measure in GDP and time to time changes in economic welfare have been widely discussed, but there has been less attention paid to the way in which the inclusions and exclusions affect international comparisons of real product per capita.

[3]See M. Gilbert and I. B. Kravis, *An International Comparison of National Products and the Purchasing Power of Currencies*, OEEC, pp. 51–57.

[4]Cf. H. S. Houthakker and L. D. Taylor, *Consumer Demand in the United States* (1966), pp. 167–72, and L. T. Wells, "Test of a Product Cycle Model of International Trade: U.S. Exports of Consumer Durables," *Quarterly Journal of Economics*, Feb. 1969.

Environment and output. One set of problems relates to the environment, including both the impact of production on the environment and thereby on welfare and the impact of the environment in its pristine state upon welfare.

Were there a way to measure them, the adverse effects of such unfavorable concomitants of high production and economic growth as high noise levels, polluted atmosphere, and congested transport arteries might be regarded as negative goods. As matters stand our conventions governing the preparation of national accounts do not provide for any deductions from GDP for environmental deterioration but do lead to additions for some but not all expenditures designed to improve the environment or to prevent further deterioration. Expenditures by government or households or by firms on capital account to reduce pollution or otherwise protect the environment are counted as additions to final product. Expenditures for these purposes by firms on current account, however, are not regarded as final product but merely increase the prices of output. This means that efforts to offset the deterioration of the environment affect the GDP estimates of different countries according to their extent and according to the transactors that carry on these efforts. Thus as between two countries one of which devoted substantial resources to attempts to combat environmental deterioration, while the other used none of its resources for this purpose, the former would, other things (resources, productivity, etc.) being equal, have a relatively smaller GDP if its environmental efforts were financed through business expenditures on current account. If, on the other hand, its efforts were conducted through governmental expenditures, the two countries would be shown to have equal GDPs.

Given the present stage of international comparisons, we make no effort in the ICP to deal with these problems. At a later point, if environmental costs continue to rise as seems likely, it may be necessary for national income statisticians to reconsider the treatment of such costs both for time to time measurements within the country and for country to country comparisons.

Differences in environmental conditions raise further questions about the comparisons of real product quite apart from the adverse effects of production. Some of these questions are relatively easy to answer. For example, a cold climate requires men to produce heat for residences and other buildings which is not necessary in a warm climate. The production of heat is an economic activity that adds to welfare and must be counted as a part of the contribution the economy is making to welfare where the heat is produced. Thus the income of a country which requires and produces heat is higher than the income of a country in a warm climate which does not require or produce heat, the production of all other products being equal in the two countries. It is equally clear, on the other hand, that added inputs or costs to attain a given product necessitated by a harsher environment do not represent more production. A potato remains a potato whether it takes one hour to produce in a rich soil in a hospitable climate or whether it takes three hours to produce in a barren soil in an unfavorable climate.

A question arises, however, into which of these two reasonably clearcut cases—the one representing added outputs necessitated by a cold climate and the other representing merely added costs—we should place the extra inputs and

costs that may be required to produce capital goods in an unfavorable environment. For example, a mountainous country may have to build largely curved highways at a steep incline, while a flat country may be able to build most of its highways straight and level. Is flat land for highways analogous to fertile land for potatoes and mountainous land analogous to barren land? In that case we should take as our unit of output one mile of highway, regardless of its inclination. Alternatively, we may argue that a mile of mountain road is regarded as more output than a mile of flat road within each country and should be counted as such in comparisons between countries. Or, to take another illustration, suppose that in a cold climate a steam power plant has to be built with walls around its boiler room and switchhouse, while in a warm climate both may be exposed to the weather. Assuming that all other characteristics are identical, shall we regard inputs and costs required for enclosed construction in the cold climate simply as added costs or shall we regard them as more output?

The answer to such questions depends in principle upon the future flow of services that each capital good would produce in each country. If we could, we would compare internationally the present value of the increases in output, ultimately in the form of consumption goods, that new capital goods would contribute in each economy. Let us imagine, for example, that production is confined to a single consumption good and to machinery designed to make possible the future output of the consumption good. Let us suppose further that we have the following information about the current output of the consumption good and its future output made possible by the machinery built during the reference year in both France and the U.S., the two countries we wish to compare:

	Consumption Good			
	Current			Future
	Price	Quantity	Value	Quantity
France	5 frcs	100	500 frcs	40
U.S.	$1	100	$100	20

One possible solution is to value future quantities at current prices in both countries. In that case, French GDP would come to 700 francs and U.S. GDP to $120 and since for current goods 5 francs has the same purchasing power as $1, the ratio of French to U.S. GDP would be 1.167. However, this would assign an unduly large weight to future goods since their present value is less than that of present goods. If we knew the time patterns of these future flows and the rate of discount in each country, we could make the necessary correction. If, for example, the average time into the future at which the future goods will become available is 10 years in both countries and the rate of discount is 15 percent in the U.S. and 10 percent in France, the present value of future goods in the two countries comes to $4.94 and 77 francs, respectively. The French–U.S. GDP comparison could then be made at U.S. relative prices for present and future goods ($1 and

$0.247, respectively) or at French relative prices (5 francs and 1.93 francs, respectively). French GDP at U.S. prices thus comes to $109.88 or 1.047 times the U.S. GDP ($104.94). When French prices are used, the U.S. GDP is 538.6 francs, the French GDP 577.2 francs, and the French–U.S. GDP ratio is 1.072.

In the real world we do not have a dated list of consumers goods that will eventually flow from new investment. An indication of the present value of these goods is, however, given by the dollar cost of the machinery in the U.S. and by the franc cost in France, that is, by the investment figures themselves. We might with some sweeping assumptions about similarities in the relative content and timing of the two bundles of future goods make the kind of comparison outlined above. For each country,

$$I = P_M Q_M = \sum_{t=1}^{t=n} \frac{P_c Q_c}{(1+r)^t}$$

where I is investment expenditure, P prices, Q quantities, M machinery, c consumer goods, r the rate of discount and t the number of years into the future when the consumer good will appear. If this equation for one country, say France, is divided by that for another, say the U.S., the ratio of the future consumption quantities $[(Q_c)_F/(Q_c)_{U.S.}]$ can be estimated since the relative investment expenditures $(I_F/I_{U.S.})$ are known, the relative consumer goods prices $[(P_c)_F/(P_c)_{U.S.}]$ may be taken from the price comparisons for current goods,[5] and rough assumptions can be made about r and about a simple average t (the latter to avoid the necessity of distributing the $P_c Q_c$'s over time).

The rather heroic assumptions required in this approximation to the theoretical ideal make it unsuitable for implementation, but it is helpful in keeping before us the notion that the appropriate criterion in capital goods comparisons is their power to produce future goods. It seems likely that we shall be able to produce somewhat firmer estimates of the price and quantity ratios for investment goods if we try to base the comparisons not on the future flows of goods themselves, but on the capacities of newly installed machinery to produce those goods. If the capacity of each new machine or other form of investment to contribute to production could be measured in terms of one dimension, our task would be relatively simple. If, for example, the contribution to future production of a steam power plant could be measured solely by the kilowatts it generates, we would have an unambiguous basis for quantity comparisons. In fact, of course, other things, such as fuel economy, labor requirements and reliability, affect the contribution of the power plant. The same is true of most other producers' durable goods.

For those kinds of producers' durables which are marketed in a variety of models, it is possible through statistical analysis to relate observed prices to the various physical characteristics that contribute directly (e.g. the horsepower of a tractor) or indirectly (e.g., weight as a guide to durability) to the value of the good in production. That is, the presumption is that the prices producers are willing to pay for these characteristics reflect their contribution to production. For goods which are available in a variety of models, we are, in the ICP, making price

[5]Although this would ignore the impact of different amounts of relative investment in the two countries on future relative prices.

comparisons between different countries in terms of goods specified for certain combinations of these characteristics. These methods, sometimes referred to as "hedonic indexes",[6] will be discussed in the ICP context elsewhere.

For other machinery and for construction, when we had no empirical evidence on expected contributions to production, we base our price comparisons on equivalent physical specifications. In the case of highways, for example, we were not able to obtain any basis for assessing the future flows of services and therefore compared prices for things with like physical specifications. The cost of a flat road of a given specification in one country was compared with the cost of a road of the same specification in another country, and likewise for a mountainous road. The effect, of course, was to treat the mountainous road as more output than a flat road; an adverse environment in this instance was in a sense requiring more production just as it would if low temperatures necessitated the provision of heat and warm clothing.

The "general" quality of output. The basic approach to the treatment of differences in quality for specific goods is to attempt to avoid such differences between goods for which price comparisons are made or to adjust prices where goods of equivalent quality cannot be found. Quality differences of this type do not, therefore, affect the output comparisons and thus need not be discussed in connection with the definition of output.

There are, however, some kinds of quality differences associated with the whole aggregate of goods rather than pertaining to specific products. For a given aggregate of goods, it is more advantageous to the population to have conveniently located, well-stocked stores with courteous and efficient sales personnel, than to have to search for supplies and to queue up for service. A similar point applies to ancillary services such as credit, delivery, right to return merchandise, and repairs and adjustments. A greater variety of goods is also to be preferred, generally, to a lesser variety.

A retail distribution system that provides all these conveniences and services is more expensive and absorbs more real resources than one which does not. The GDP of each country includes the value of such services to the extent that they are rendered. It is important to note, however, that our method of international comparisons does not attempt to measure international differences in their provision. We simply compare the extent to which each economy delivered meat and potatoes and shoes and stockings and the like to its residents without regard to the extent or nature of the accompanying services.

It is very difficult to judge the direction of the bias that results from the omission of these general quality factors. There are, for example, some reasons to believe that American output would be higher relative to that of other countries were it possible to measure and include these factors in our comparisons. Variety, availability of goods, convenience of shopping, and the provision of services including credit sales all appear to favor this point of view. On the other hand, Americans tend to buy in relatively large quantities, particularly

[6]Z. Griliches, "Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change," *The Price Statistics of the Federal Government*, National Bureau of Economic Research (1961), and I. B. Kravis and R. E. Lipsey, *Price Competitiveness in World Trade*, Ch. V (1971).

with respect to foods and household supplies, with the consequence that a part of the storage function is transferred from retail establishments to households. In addition, a larger number of sales in small quantities more typical of other countries can be regarded as the provision of more distributive services for the same quantity of final product delivered to consumers. Thus even if it be granted that U.S. shops are more convenient and provide more services in some respects than those of other countries, there are at least some offsetting factors.

Our decision to ignore all these differences and to define our commodities primarily in physical terms has implications for the form in which we cast our price comparisons. It means that we should obtain for each specification the average price paid per unit for the average-size transaction in each country. If, for example, a certain specification of soap is customarily purchased in packages of three in one country and of six in another, we should obtain the price per cake for packages of three in the first country and the price per cake for packages of six in the second. This treatment, by the way, has the merit of producing a price for each country which is consistent with the expenditure figure in the national accounts.

Other aspects of the problem of determining national average prices will be taken up after a discussion of the problem of the selection of representative items for price comparisons.

The Comparison of Prices

The quantity comparisons sought in the ICP could in principle be achieved through direct quantity comparisons for each detailed subcomponent of GDP (milk, tractors, etc.); these quantity ratios could then be aggregated by the use of expenditure weights. Each quantity ratio, whether referring to a subcomponent or an aggregate, could, by division into the corresponding expenditure ratio, be used to derive a corresponding price ratio. That is,

$$\frac{P_A Q_A}{P_B Q_B} \div \frac{Q_A}{Q_B} = \frac{P_A}{P_B}$$

where A and B are countries and the P's and Q's refer to prices and quantities.

Quantity comparisons are, however, difficult to make for many categories of products. Apparel, for example, is so heterogeneous that quantity data for each type and quality of product are often difficult to obtain. Even if quantity information about some variants is available in two countries, the quantity ratio may be expected to exhibit wide dispersion in the case of categories composed of such varied kinds of products.

The alternative approach is to make price comparisons and to derive the quantity ratios from them and the expenditure ratios. This approach has the advantage that price ratios for individual products are easier to obtain and are subject to less dispersion and smaller sampling error than the corresponding quantity ratios. Primary reliance is therefore being placed on the price comparison approach in the ICP, although quantity comparisons are also being used, mainly as a check on the results of the price comparisons.

Sampling Principles

For the purpose of making the comparisons, GDP has been subdivided into approximately 160 classifications which we refer to as detailed categories. For each of these categories, it is necessary to select a number of representative goods for which prices for identical or equivalent specifications can be found in both countries of each binary comparison. Before describing what has actually been done along these lines, it may be useful to consider what we would do if our resources were unlimited and our knowledge complete. It is convenient to think of these problems at least initially in terms of binary comparisons (i.e., involving only two countries), although our ultimate objective is a set of consistent multilateral comparisons (i.e., involving many countries).

One way to think about the problem is to begin with the population of the final purchases, unit by unit, of commodities and services in each of two countries in a binary comparison. Each population of transactions may be divided into those which are for commodities and services that are common to the other country and those that are for things which are not included in the other country's set. It might be possible one way or another to establish equivalences between things in the non-overlapping sets of the two countries, but for present purposes we will ignore this possibility and concentrate on the items in the overlapping set.

In principle, then, an international price comparison would be based on a random sample of the price relatives (the ratio of one country's price to the other's price)[7] of the commodities and services found in the overlapping set. Of course, the character of such a random sample and of the population of overlapping items which was being sampled would be unambiguous only if the frequencies of the purchases of identical products were the same in the two countries. Actually, this is extremely unlikely and the probability is that some items in the overlapping set are purchased more frequently in one country than in the other. Let us put this problem aside for the moment, however, and define the overlapping set to include not only the identical items in the two countries but also the identical numbers of each item. We would in effect take the highest common multiple for each item and put the excess into the non-overlapping set.[8]

In a random sample of the population of identical items appearing with identical frequencies, each transaction in a final product would have an equal chance of being represented in the sample. Even this approach would have the disadvantage that it would lead to an oversampling, from the standpoint of a value aggregate such as GDP, of items of small value with numerous transactions relative to those of high value with few transactions. This would not bias the estimated purchasing power ratios between currencies unless relative international prices varied systematically with the size (in value terms) of the transaction unit in which different goods are typically exchanged. However, it is more likely than not that relative prices do vary in this way, since one would expect

[7]That is, P_A/P_B, where P is the price of a given specification of a good and A and B are countries.

[8]This is similar to Keynes' "highest common factor". See J. M. Keynes, *A Treatise on Money*, vol. 1, p. 108 (Macmillan, 1950).

durables which tend to come in units that carry larger price tags than nondurables to be relatively cheaper in higher income countries.

A more appropriate sampling frame would involve the substitution of values for physical quantities so that the population of final products being considered would be a population of dollars or pounds sterling or kroners worth of transactions. The difficulty here is that the evaluation of the overlapping set of commodities would depend upon which country's prices were used to value them. A valuation in the currency of either country—so as to obtain a sampling of the distribution of say dollars' worth or kroners' worth or other currency's worth of final product—would be an improvement over the sampling of the common list of physical items, although the result would probably be to produce higher relative prices for the country whose prices were used than would be produced if the other country's prices were used. This effect can be expected due to the tendency for relative prices to be inversely correlated with relative quantities, referred to above. The use of one country's prices would tend systematically to assign lower transactions values to those kinds of goods that were most important in that country and therefore to diminish the sampling ratio for such goods.

A preferable way of dealing with the difference between the value and physical unit distributions would be to sample the physical unit distribution and then to weight the different price relatives according to their expenditure weights for one country or the other or a combination of both countries.

In the real world it would be very difficult to approximate any such ideal scheme of random sampling. For one thing, an existing stratification of the transactions is forced upon us. We are in reality not confronted with a list of individual transactions but with a classification of final expenditures divided into commodity groups which, while differing in detail are generally similar from one country to another. There is first the division into government, households, and capital formation and within each of these sectors there are fairly familiar subdivisions (e.g., food, clothing, etc., in consumption). While the widespread adoption of such a classification attests to its utility, there is little reason to believe that it is optimal from the point of view of international price comparisons. Some of the common classifications, such as dairy products, can be expected to be composed of items for which the price relatives will be fairly uniform from one specification to another, while others, such as household furnishings (including furniture, household textiles, and household appliances), may contain price relatives that vary very widely. We do of course have some control over the commodity classification by our ability to combine groups or, what is generally more helpful but also often more difficult, to subdivide them so as to obtain categories which are more likely to have homogeneous price relatives in them.

Indeed, one suggestion that has been made is to use the dispersion of price relatives as a criterion for the classification of items into commodity groups.[9] The idea is to choose among alternative classification systems the one which minimizes the variance of price relatives within categories relative to the variance between categories. This is an attractive objective since small dispersions justify the use of unweighted averages within categories, a practice which is sometimes

[9]See T. Mizoguchi, "An Application of Variance Analysis for the International Comparison of Price Levels," Hitotsubashi University (processed), May 1969.

made necessary by the lack of data for weighting and which is always convenient.[10] However, at the present early stage of international price comparison work, we have little choice in choosing a sample but to start with some fixed classification, in the selection of which the expected dispersion of price relatives can only play a relatively small part. As experience accumulates, it should be possible to modify the classification so as to minimize further the dispersion of price relatives within categories. It can be expected that progress along these lines will be easier in binary than in multi-country comparisons, particularly since in the latter the maintenance of identical detailed categories merits higher priority than the reduction of within-category price dispersion.

A more feasible procedure is to start with the basic classification used by most countries, to modify it with some subdivisions designed to improve homogeneity and to cope with the remaining problems of heterogeneity within classifications by increasing the size of the sample within the more heterogeneous categories. Optimally, the sampling rate within each category would be proportional to the standard deviation of the price relatives in the category.[11]

Sampling Practice

Turning now to our practice, we have, in the light of the theoretical and practical problems facing us, worked along three lines in choosing specifications:

(1) The classification basically follows traditional final product classification lines, but it was modified with a view towards reducing the dispersion of international price relatives within categories. As already noted, we really do not know even at this stage what the dispersion of price relatives will be in different categories, and we had to base this work on *a priori* expectations. For example, nondurable household goods (SNA Category 4.5.1[12]) was broken down for our purposes into paper products, cleaning supplies, and other items in the expectation that the price relatives for nondurable household goods would have substantially larger dispersion than the price relatives for other classifications and that this would be less true of the subcategories.

(2) The target numbers of specifications for the individual categories were determined in the first instance on the basis of the relative importance of the categories in the GDP's of five or six countries for which information was available in the early planning stage. The roughly proportional sampling ratios were then modified in the light of the expected degree of dispersion of price relatives within the categories. For example, only one specification was provided for categories like eggs and coffee while five were called for in the case of a category like men's and boys' hosiery, underwear and nightwear.

(3) Two principles were used in the selection of specific items within the detailed categories:

[10]The convenience arises from the fact that unweighted averages produce single price and quantity comparisons, rather than different ones for each set of weights. Single averages can, it is true, be produced by formulas that take into account the weights of more than one country, but they are less appropriate for indirect comparisons which, for example, derive the Country *A*/Country *B* comparison from the product of the Country *A*/Country *C* and Country *C*/Country *B* comparisons. (See "Methods . . ." paper, p. 25.)

[11]See L. Kish, *Survey Sampling*, p. 92 (1965).

[12]See U.N., *A System of National Accounts*, p. 107.

(a) The criterion of "concentrated selection", that is, the selection of the goods with the largest expenditure weights, was adopted. The advantage of this rule is that it produces a large coverage of the expenditures within each category at a low cost, and thus diminishes the likelihood of sampling error attributable to omitted items.

On the other hand, concentrated sampling has significant disadvantages for our purposes. In the first place, it yields an unambiguous rule for the selection of items only if applied to the expenditures of one country. When used in this fashion, however, it is likely to produce a sample of items some of which will fall outside the overlapping sets referred to above. Even if applied from the standpoint of one country to the items within the overlapping set, concentrated sampling will bias the price comparisons so as to produce lower relative prices for the country whose expenditures are used as the basis for the selection of items.[13] The reason is, once again, the inverse correlation between price relatives and quantity relatives.

The second disadvantage of concentrated selection is that it is likely to lead to an under-representation of the items of low importance within each group.[14] If the price relatives are very different for low- and high-volume items, this will bias the results, although it is more difficult to predict the direction of this bias in international comparisons.

(b) Each specification chosen had to be important in the consumption of each country, or important in one country and at least in common use in the other. The idea is to avoid the selection of items which while they can be found in a given country will be so uncommon as to provide an unrepresentative basis for price comparisons within the category in which they fall. This means that each specification should be typical for the category in each of the countries with respect to volume of sales, source of supply (domestic versus foreign), and any other factors that affect relative price formation. In the case of the U.S., for example, goods that can be readily found in stores that cater to mass markets, such as mail-order houses, supermarkets or department stores, can be taken as satisfying this criterion.

Matching Qualities of Goods

For the most part, what was referred to earlier as the "specific" quality problem is dealt with by attempting to ensure that prices are compared for items that are physically identical in different countries. In a number of cases, even the same model of a given manufacturer can be priced in two or more countries (e.g., Colgate toothpaste, Caterpillar tractors). There are other instances in which physical identity is not present, but the differences between the two products appear to be relatively unimportant or offsetting. Thus in the case of most foods, for example, it has been assumed that items are alike in quality in the several countries unless there is some specific indication to the contrary. For meats, for example, it seems clear that there are substantial differences in the quality of meat in some countries where high income levels prevail and a highly developed animal

[13]See D. Brady and A. Hurwitz, "Measuring Comparative Purchasing Power", *Studies in Income and Wealth*, vol. 20 (1957).

[14]Brady and Hurwitz, op. cit.

husbandry industry has been developed to cater to a taste for fine meat. In other countries, where income levels are lower and meat consumption less important, the average quality is often lower. Fortunately, grading standards for meat are well developed in countries in which meat is important and these provide a basis for selecting the appropriate quality level for price comparisons with other countries. The problem posed by the different ways of butchering carcases in various countries is being dealt with by obtaining the composite average retail price per pound for each type of animal.

We have tried to use physical identity even in cases in which the methods of production are very different. With respect to men's suits, for example, machine made garments are most common in the U.S. and are cheaper than tailor-made garments, while in India the opposite is true. We have in this case sought to establish for each Indian tailored garment a U.S. machine-made garment equivalent in quality. This effort involved an exchange of samples and consultations with cloth manufacturers in both countries, tailors in India, and clothing manufacturers in the U.S.

In a few kinds of situations, equivalence has been established on the basis of similarity of use rather than physical identity. In some of these cases, products serving the same use or function may be produced most economically in different forms in different situations. For example, paper containers are cheaper for the distribution of cream in the U.S. where high wages make the collection and cleaning of used bottles expensive and bottles are cheaper for distribution in Europe where paper products are costly. In these circumstances, cartons and bottles containing the same quantities and qualities of cream are being taken as equivalent and directly comparable.

There are other instances of things that are not physically identical but clearly serve the same need or use. For example, in the United States 120-volt lightbulbs are commonly used, whereas in Europe 220-volt bulbs are common. It appears that there would be little or no difference in the cost of production were the two types of bulbs produced under similar conditions in the same country and since there is no difference in the utility afforded by them, they have been treated as equivalent products.

Sometimes there are cost or price differences, but they are found to be in the opposite directions in the two countries for two products serving a similar purpose or need. In the case of rice, for example, the Japanese prefer short grain rices and these tend to be more expensive than the infrequently encountered long grain rice, while in the U.S. the price relationship is just reversed. We have accepted the Japanese view that this difference is a matter of taste and are comparing the price of long grain rice in the U.S. with the price of short grain rice in Japan.

Special treatment in the form of "hedonic" indexes, referred to earlier, has been applied to a number of durable goods. Durables typically come in a variety of models, each offering the purchaser a different mix of the cluster of characteristics, such as power, capacity (perhaps measured by size) and comfort or durability (for which weight is sometimes a proxy variable). Of course, where the entire markets of two countries consist of identical models produced by the same manufacturers, direct price comparisons of the traditional type can be made.

This is true also when there are some unique models in each country as long as the overlapping models can be regarded as representative of each country's market. Where these conditions are not met, hedonic indexes have advantages over the older method of trying to match different models for each pair of countries. The traditional method of attempting to match products often has to be based on a rough assessment of the relative importance of the key variables affecting the relative values of different models.

It should be noted that in this approach "quality" has to be defined largely in terms of quantifiable characteristics with the possible addition of one or more other features (e.g., wheel as distinct from crawler tractors) on an "either or" basis through the use of dummy variables. It is possible that the characteristics such as weight which may serve as proxies for the desired qualities may bear a different relation to them in different countries. For example, within each of two countries heavier tractors may be more durable than lighter ones and each country's regression equation will reflect this by producing a higher price for the heavier tractor; but if at each weight the tractors of one country are superior with respect to durability to those of the other country because of superior workmanship, the regression equations will miss this. Of course, differences like this would be hard to catch in the traditional method, too, and if there were a way to measure them the results of the hedonic approach could be adjusted accordingly.

Problems of Obtaining National Averages

As already noted, the SNA framework makes it logical to seek for each specification the national average price paid by final purchasers, since this is the price which when multiplied by quantity will yield the expenditure entry in the national accounts.

There appears to be a great deal of difference among the countries in the degree of dispersion of prices for a given specification around the national average. In some countries, such as Hungary and the U.K., dispersions seem to be very small. In others, such as the U.S. and India, there appears to be a very substantial dispersion of prices and the problem of computing national average prices is a formidable one.

In the case of the U.S. the prices of some goods vary substantially according to the type of outlet, with discount stores and supermarkets charging lower prices than other shops. In general, the Bureau of Labor Statistics, the official price collection agency of the United States, has carefully designed its sample of outlets in the light of the relative importance of the several types in the sales of each commodity. The BLS itself computes and publishes average national retail prices only for food items. For non-foods, the BLS price collection operation is designed to measure time to time changes in price and does not lend itself readily to the computation of national average prices. The time to time needs of the BLS are met perfectly well, for example when different variants of a particular specification are priced in different outlets, as long as the specification priced from time to time in a given outlet is unchanged. The regular preparation of national average prices is limited largely to foods regularly priced for the consumer price index. Because in this initial stage of the ICP the U.S. was being

used as the base country for nine other countries, substantial numbers of U.S. national average prices—running into the hundreds—were required. Ways therefore had to be found to develop national average prices through short-cut methods.

For the nonfoods, the BLS gave us tabulations of prices for the selected specifications in five large cities of the U.S. for October 1970. There were from 15 to 25 quotations per specification. The average price for each specification in each city was adjusted to an estimate for the national average price based on place to place price relationships computed from data gathered for the *City Workers' Family Budget.*[15] For example, in Autumn 1966, the date of reference for the *City Workers' Family Budget*, the bedsheet priced for the budget was 2.3 percent higher in New York City than the U.S. average price. This relationship was used to estimate a national average price for October 1970 from the New York price for that date for the specifications of bedsheets included in our sample. National average prices were estimated from the bedsheet prices for the other cities by a similar procedure. The five estimates of the national average price obtained from the five cities were then averaged together without weighting since there was no reason to believe that an estimate based on any one of the cities had a claim to more weight than another. The national average obtained was then adjusted to calendar 1970 and to other years by the most appropriate BLS item index.

For other items the BLS prepared special tabulations of prices collected for the *City Workers' Family Budget* in 1966 for New York and Chicago. These, too, were adjusted to national average prices and to the desired dates by similar methods to those described above.

For items not in the regular lists of specifications collected by the BLS, special collection procedures had to be arranged. For a number of foodstuffs several major supermarket chains were contacted and records of their monthly average prices in a number of different cities in which they operate were obtained. Special inquiries were also made in the case of a few items which are important for developing countries, but which would be difficult to gather through the normal price collection machinery of the BLS. Special written and telephone inquiries were made to obtain a nationwide sample of firewood prices, the identity of the sellers being determined from the yellow pages of telephone books in different cities. For comparisons of suits and overcoats with countries in which such clothing is typically tailor-made, cloth samples were sent to a number of manufacturers who were asked to provide the retail prices of specified garments and the share of labor and material costs in the prices they obtained from their customers. Prices were also obtained directly from manufacturers for items slated for regression analysis, including fans, air conditioners, TVs, refrigerators, and washing machines.

Prices for a substantial number of additional specifications were gathered by the Bureau of Labor Statistics. The national price collection machinery was used to gather prices for more than a hundred specifications, and the Philadelphia office of the Bureau of Labor Statistics was used to obtain prices for well over 200

[15] *City Workers' Family Budget* (*CWFB*), *Pricing Procedures, Specifications, and Average Prices; Autumn 1966*, Bulletin No. 1570-3, U.S. Department of Labor, Bureau of Labor Statistics.

additional items. The latter group of prices had to be adjusted to the national average by methods described earlier.

In India the problem of obtaining national average prices took a different character, and the methods used to cope with them were accordingly different. In this large and varied country, not only do prices for the same thing differ from one area to another, but in many cases the specifications of the items consumed in large volume also vary substantially. This is the case, for example, for rice which is found in a large variety of types which differ from one part of India to another. Indeed, the general content of living in the rural areas of India is so different from that of a western country such as the United States that a direct comparison would be very difficult.

The solution of the problem is to make the price comparisons with the U.S. in three successive stages. First, prices in metropolitan India (Bombay, Calcutta, New Delhi, and Madras) will be compared directly with U.S. prices. Then, a second stage of comparisons will be made between prices in the metropolitan areas and other urban India, and a third set of comparisons between other urban India and rural India. Each stage will be based on its own sample of commodities chosen from the overlapping set of specifications available in the relevant pair of areas. The "other urban"-rural comparison, which has already been carried out, has been executed separately within each of the eighteen states; since specifications differ less within each state than across India as a whole, the problems of finding matching specifications and of controlling quality have been reduced by this means.

The Treatment of Consumption of Own Production

A special problem in the concept of national average price arises in connection with the treatment of consumption of own production. Like many price problems, it is directly related to the concept of national product.

Under conditions of optimal resource allocation, goods that absorb more resources represent more output than goods that absorb less resources. Potatoes consumed in the city count for more output than potatoes consumed on the farm. This valuation, based on the conditions of static equilibrium, is embedded in national income accounting practice, and, to some extent, it is carried over into the time to time measurement of changes in income. Thus, in national accounts statistics as usually prepared, a shift from farm to urban consumption, with farm and urban prices constant, raises real product. If we try to match this treatment in international comparisons, we will treat own consumption of potatoes and purchased consumption of potatoes as separate products. This implies pricing the former at producers' prices and the latter at retail prices.

This method leads to a lower estimate of the relative product for a country with relatively high own-consumption and with a relatively large spread between the producers' and retail prices. The illustration may bring out the way the arithmetic works.

Both countries consume 110 of potatoes but with very different distributions between own consumption and purchased consumption. The quantity index with the urbanized country's weights is $(\Sigma P_U Q_R / \Sigma P_U Q_U) = (600/1050) = 57$. The quantity index with the rural country's weights is $(\Sigma P_R Q_R / \Sigma P_R Q_U) = (280/820)$

	Urbanized country			Rural country			Extensions	
	P_U	Q_U	P_UQ_U	P_R	Q_R	P_RQ_R	P_UQ_R	P_RQ_U
Own consumption of potatoes	5	10	50	2	100	200	500	20
Purchased potatoes	10	100	1,000	8	10	80	100	800
Total		110	1,050		110	280	600	820

= 34. If the difference in price spread is eliminated so that the retail (purchased) price in the rural country is 4 (i.e., two times the own consumption price as in the urbanized country), both indexes come to 57. (Or, if the common spread is four times the own consumption price, both indexes come to 34.)

Whether it is right or not to use a method that results in counting the same quantity of potatoes as less output because they are consumed on the farm depends on the reasons for the different distribution of populations between farm and city. In general, a population living close to its points of production will require less transport inputs and will, all other things being equal, enjoy more final product than a population dwelling at a greater average distance from its points of production. Distance involves a cost, and a greater need to overcome it should not be allowed to count as more output any more than the greater need to wrest production from a less fertile soil. Someone might wish to argue that urban dwelling (at a distance from production) is a result of choice rather than necessity and that the greater costs entailed in this preference constitute a contribution to welfare. There may be an element of truth in this view; the attraction of cities all over the world seems to be powerful and only partly explicable in terms of greater economic opportunity. We cannot, however, gauge the relative roles of choice and necessity in urban concentration. We are not trying to assess the relative utilities and disutilities involved in urban dwelling. We do not attempt to deduct the disutilities entailed in urban life (pollution, commuting time, etc.), and we should eschew techniques that implicitly ascribe greater utilities in urban living. Counting an urban potato as more product than a rural potato (owing to the costs of transport and trade margins) would be to ascribe more welfare to city than to rural dwelling. It seems preferable to regard a potato as a potato for international comparison purposes.

The way to achieve this is to combine own production and purchased output for each product into a single category. This still leaves open the question of how the national average price will be determined, particularly how own consumption will be valued for the purpose of estimating the national average price. There are two possibilities for the valuation of own consumption. One is to value it at retail; the other is to value it at producers' prices. In the latter case the national price is the weighted average of producers' and retail prices, using consumption weights. In either case, once the national average price is determined, both methods treat all units consumed so that they make an equal contribution to each country's relative product; a potato is a potato for comparison purposes, whether consumed on the farm or in the city. What is different is that the relative importance of the product—that is, the price weight assigned to the quantity—

will be greater when retail prices are used than when weighted average prices are used. In an international comparison the quantity ratio for potatoes will be the same whichever prices are used, since all potatoes are treated as a single category, but potatoes will have a bigger weight for purposes of calculating the overall GDP ratio if the former are chosen. There is, however, a clear case for using weighted average prices. Firstly, these are the prices that truly reflect the average resource input in each country. Secondly, they are the prices that are in each country's expenditure data (if the SNA is followed), and they are therefore the prices that will produce price ratios consistent with the appropriate quantity and expenditure ratios.