NBER WORKING PAPER SERIES

THE INTERNATIONAL COMPARISON PROGRAM: CURRENT STATUS AND PROBLEMS

Irving B. Kravis

Robert E. Lipsey

Working Paper No. 3304

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 1990

NBER Working Paper #3304
March 1990

THE INTERNATIONAL COMPARISON PROGRAM: CURRENT STATUS AND PROBLEMS

ABSTRACT

This paper reviews the International Comparison Program (ICP), a worldwide effort to produce international comparisons of real GDP and its components and purchasing power parities of currencies (PPPs). The robustness of results and future work are considered.

A generous estimate of margins of uncertainty in the benchmark estimates might be 20-25 per cent for low-income countries and 7 per cent for high-income countries. The errors in extrapolations to countries not covered by the surveys could go as high as 30-35 per cent. That is still a small range of error compared to that stemming from the use of exchange rates to convert own-currency to common currency measures of output. Furthermore, exchange rate conversions are even more sensitive to methodology than PPP conversions. The notion that exchange comparisons rest on a simple and transparent procedure using standard market data is illusory.

The future of ICP measures seems assured in Europe, particularly in the European Community. The prospects for systematic worldwide comparisons do not look as bright. A renewed effort by the United Nations Statistical Office and the World Bank would be needed to maintain an ICP with comprehensive coverage and comparable methods in all major regions.

Irving B. Kravis
University of Pennsylvania
Department of Economics
Philadelphia, PA 19104
and NBER
269 Mercer Street, 8th Floor
New York, NY 10003

Robert E. Lipsey
Queens College and The
Graduate Center, CUNY
Flushing, NY 11367
and NBER
269 Mercer Street, 8th Floor
New York, NY 10003

Introduction

A world wide program to produce international comparisons of real GDP[1] and its components and the purchasing power parities (PPPs) of currencies has been under way for 20 years. In this paper we review the methodology of this work, the International Comparison Program (ICP), and describe its present status. (The effort was referred to as the International Comparison Project in its earlier stages.) We follow with a discussion of the robustness of the results, some problems that have arisen in the extension and continuation of the ICP, and a look to its future.

The predominant method of meeting the need for comparative data on real GDP and related macrovariables is to convert own-currency value aggregates to a numeraire currency, usually the dollar, via exchange rates (as, for example, to obtain world expenditures on energy). Exchange rate conversion is still the common practice despite clear evidence that exchange rates fail to reflect the relative purchasing power of currencies, sometimes being off by a factor of 3 or more, even for output as a whole and still more for individual products.

An outline of ICP methods[2]

The ICP comparisons relate to Gross Domestic Product (GDP) and its components as defined in the U.N. System of National Accounts (U.N. Statistical Office, 1968). The concept of GDP is intended to establish a production boundary marking off economic activity, which produces satisfaction-yielding goods and services, from other human activities. In this context, production is generally regarded as a measure of income, although there are circumstances in which it seems appropriate to distinguish between the two concepts.

The ICP approaches the international comparison of income through price comparisons for about 150 categories ("basic headings") of final expenditures on GDP[3]. Prices are usually compared for at least several specifications of goods in each category. Because the items priced in each country must be representative of the goods commonly found in the domestic market, a common list of price compared goods in all countries is precluded. There are two ways in which category PPPs are calculated despite incompletely overlapping price comparisons for the included countries. In both, the missing prices are inferred from their relationship in all the other countries to prices of items that are available in the given country.[4]

The quantity comparisons for each basic heading are obtained by dividing the PPP into the expenditure ratio. That is, for a given basic heading,

$$\frac{Q_j}{Q_b} = \frac{E_j}{E_b} \div \frac{PPP_j}{PPP_b}$$ where j and b are countries, Q's are physical quantities, E's expenditures in own currency.

The PPPs and the quantity indexes for the basic headings derived by these methods are then aggregated to form PPPs and quantity indexes for summary headings (e.g. "meat," which includes fresh beef and five other basic headings), and for larger aggregates up to GDP. The method of aggregation involves the use of a set of average international prices (π's), one for each basic heading. The average price is calculated from the category PPP for each country weighted by the quantity absorbed in that country. To make the PPPs of the different countries commensurate, each is divided by the overall PPP (i.e., the PPP for GDP) for the country. Thus,

$$\pi_{ij} = \Sigma \frac{PPP_{ij}}{PPP_j} \cdot w_{ij}$$

where i is one of m categories or basic headings and w the share in world consumption.

The GDP PPP is the ratio of the country GDP at own currency prices to its GDP valued at international prices:

$$PPP_j = \frac{\sum^{m} PPP_{ij}\, Q_{ij}}{\sum^{m} \pi_i\, Q_{ij}}$$

Thus the PPP's for GDP (or other aggregates) need the π's for their calculation, and the π's need the PPP's for their calculation. The solution, suggested by R.C. Geary and amplified by S.H. Khamis, is to rely on a set of equations in which all the PPPs and π's are simultaneously determined. To obtain GDP or other aggregates the quantity in each basic heading is

multiplied by the international price, and the products obtained are summed over the appropriate headings.

A number of other index number formulas were considered. Geary-Khamis was selected both for its statistical properties and for its ready economic interpretation. The statistical properties include base country invariance[5], transitivity[6], and matrix consistency[7]. From an economic standpoint, it matches the underlying point of departure of the PPP comparisons - viz., that there is a price level for each country which is an average of the different price levels of its GDP components. The use of cross country averages of these relative prices to value the quantities in each country's GDP provides a common measuring rod for the GDPs of the included countries. This price-times-quantity-equals-expenditure feature of the Geary-Khamis formula fits well with national income accounting concepts.

A persistent finding in all phases of the International Comparison Program, which now covers 80 countries, is that the purchasing power of the currencies of low income countries is much greater than that indicated by exchange rates. In 1980, for example, the price level of the developing countries of Asia included in the Phase IV study (8 countries) was only half of that of the U.S. while for the developing countries of Central and South America, the price level was less than two-thirds that of the U.S. This means that the real income per capita of the Asian countries was twice that suggested by exchange rate conversions and that of the Central and South American countries was half again as much as the exchange rate conversions indicated. It is important to bear the size of these differences in mind when considering the margins of uncertainty that are attached to the ICP comparisons.

Robustness

Some errors in real income comparisons originate in the estimates of GDP and the national accounts of various countries. The accuracy and comprehensiveness of the accounts vary from one country to another. Countries without well developed economic statistics can produce national accounts only with large margins of error. Also some parts of the accounts, like the measurement of subsistence income or of depreciation, are particularly vulnerable to differences in treatment by the national statistical authorities of different countries. Another major source of error is introduced when the data are put on a per capita basis; for some developing countries population estimates are subject to very wide margins of error.

These uncertainties in the estimates of each country's per capita national income pose the same problems for exchange rate conversions as for PPP conversions. The seemingly unequivocal character of exchange-rate-based comparisons is an illusion; it involves the conversion of what may be quite inconsistent measures into equally inconsistent estimates in the numeraire currency.

Price comparisons

There are, however, other sources of uncertainty in the PPP conversions, and we try here to evaluate them. There is first and most basically the problem of matching qualities in the price comparisons. The care taken to get these comparisons right varies from one phase of the ICP to another, and sometimes within a given phase from region to region or country to country. The comparisons involving the countries of the European Communities (EC) meet these problems in a particularly thorough and careful way. The price comparisons involving the U.S. for the Phase IV study were less than optimal

because the decision of the U.S. to participate was taken late and prices collected for an earlier period had to be used (after adjustment to the reference date). It would be very difficult to measure the possible errors arising from incomparable qualities. The fact that so many individual prices of many items are compared (300 to 700 items per country) should help to reduce the sampling error and the fact that the price comparisons are distributed over the entire gamut of final products, may diminish the likelihood of error from biased selection among types of goods. However, it is difficult to know what biases are introduced by the need for comparability of specifications among countries or the problem of "nonresponse" arising from the unavailability of data.

Sampling variability as well as errors in matching is also difficult to measure. An experiment reported in the Phase I study (Kravis, Kenessey, Heston, and Summers, 1975, p. 77f.) suggests that the sampling variability is generally in the 5 to 7 percent range at the 0.95 confidence level, with high income countries at the lower end and low income countries at the higher end, with one case of a sampling error of nearly 10 percent. (See also Kravis, Heston, and Summers, 1982, p. 97.)

Variations in the results of the comparisons could also arise from differences in aggregation methods and from differences in the treatments of certain problem categories. Our method of assessing these uncertainties is the very crude one of examining the variations in results produced by the alternative methodologies using data of past comparisons, often those of 1975 because that study was the most recent offering much relevant material for this purpose.

Aggregation methods

The use of the Geary-Khamis formula has not gone completely unchallenged (Isenman, 1980; Drechsler, 1988), despite a favorable verdict in an influential report by Peter Hill (1981), commissioned by the Statistical Office of the European Community and the Economic Commission for Europe.

The main objection to the Geary-Khamis formula is that it produces the Gerschenkron effect (also referred to as Bortkiewicz's Law, or the own-price effect): Because the ICP prices are averages weighted by the quantities absorbed in each country, the world average price structure used in the ICP version of Geary-Khamis is heavily influenced by the U.S. and the other high income countries. Given the usual negative correlation between quantities and prices, the real GDPs of the low income countries will tend to be higher than if a set of prices more "characteristic" of them was used.

However, there is some evidence that the impact of different price regimes on Geary-Khamis estimates of real GDP is modest compared to the difference between PPP conversions and exchange rate conversions. When the Geary-Khamis index for the Phase III countries was recalculated using equal weights for poor and rich countries in obtaining world average prices, the real GDPs per capita for the eight poorest countries (real GDPs per capita 15 per cent or less of the U.S. level) were smaller only by an average of 11 per cent (a range of 9-13 per cent for the individual countries). Exchange rate converted per capita GDPs were 60% lower on average than the PPP converted figures, and one-third lower in the case of the smallest difference. (Kravis, 1984, p. 33.)

The fact that objections are still raised to the Geary-Khamis formula leads us to report the difference that it would make if the chief alternative

aggregation formula, the EKS, were used. The EKS is based on Fisher indexes which are regarded favorably on the grounds of characteristicity.[8] The Fisher indexes are not transitive and few would favor them for multilateral comparisons. The EKS (applied here to aggregate the PPPs for the basic headings), is transitive and minimizes the squared log deviations from the Fisher indexes. However, the EKS does not produce matrix consistency (its failure is in additivity; i.e., the sum of the basic headings does not produce the same GDP as the formula), and further transformations have been sought, but no widely acceptable one has been advanced.[9]

For the (nine) high income countries in the Phase III study the Geary-Khamis and EKS formulas produced nearly identical results, except for Japan for which Geary-Khamis was 5% higher. The spread was larger for middle income countries and greatest for low income countries, averaging around 16 per cent for the lowest fourth and reaching as high as 19 per cent.[10]

Treatment of problem categories

An aspect of the methodology that bears upon the robustness of the results is the treatment of certain categories for which there is little theoretical guidance for the choice between alternative methods.

Methodological differences arise both because of efforts to improve the procedures from one Phase to another, and from different choices on issues for which there is no clear right or wrong. In the first three Phases of the ICP (1970, 1973, and 1975) the same team at the University of Pennsylvania, in close collaboration with the U.N. Statistical Office (UNSO), produced the comparisons and the differences in methods resulted from efforts to refine and improve the treatment of some difficult categories. When in the later phases of the ICP, the design and collection of the ICP was organized on a regional

basis (more on regionalization presently), differences arose even within a given phase.

Aside from differences in results owing to regionalization per se, the differences related mainly to the treatment of the net foreign balance and about a dozen "comparison-resistant" service categories for which output data are difficult to devise and for which domestic national accounts generally use input data to measure output (health, education, and government services).

Comparison resistant services. The proportion of GDP spent on services ranges from less than 20 per cent in low income countries, to 50 per cent in high income countries[11]. Services that lend themselves to price comparison make up more than half of the total. The others -- the comparison resistant services -- are comprised of the services of health care professionals, teachers, and government employees. They are "comparison resistant" because it is difficult to find markets on which units of these services are sold; consequently, there are cases in which no market price paid by final purchasers is available. In domestic national accounting the absence of quantity indicators of output has been met by using changes in inputs as measures of changes in output, and a similar strategy has been followed in the ICP for the most part. In most but not all parts of the ICP, it was assumed that the productivity of workers with similar qualifications was the same in different countries. In the education sector, for example, it was assumed that a teacher in India with thirteen or fourteen years of education produced the same amount of output as a teacher in the U.K. with thirteen or fourteen years of education. Exceptions to this equal productivity assumption were made in Phase III and for the portion of Phase IV comparisons involving Austria and the three eastern European countries.

Another difference in the treatment of comparison resistance services was that allowances were made in Phases I to III, but not in later phases, for capital inputs in the health and government service categories. The capital data for the adjustment were very rough and it is understandable that the international organizations that produced Phases IV and V of the ICP abstained from their use, although the result of omitting them is surely an understatement of output in high-income countries, since both services and government services must be much more capital-intensive in high income countries than in low income countries.

Still another difference in comparing these difficult services was the use in Phase III of the number of students, in addition to the input of teachers and of capital. Although the change may have been introduced with the idea of adding a measure of the output of education, because students' time is an input into the production of learning, learning itself being the output, the procedure of Phase III represents an enlargement of the coverage of inputs to include students' time input. However, the procedure does not take foregone earnings into account or, in effect, treats foregone earnings per student as identical among countries. It therefore is likely to understate education input in high income countries relative to low income countries.

How much difference do these variations in the treatment of comparison resistant services make for the quantity comparisons for the sectors involved and for GDP? The answers are: not much for the developed countries either at the sector level or for GDP and quite a lot for low income countries at the sector level, and modest amounts at the GDP level. What is probably a maximum estimate of the amount involved in GDP is provided by doing the GDP comparison

on the assumption that the PPPs for comparison resistant services are the same as the PPPs for priced services. (Actually, in the Phase III report the PPPs for comparison resistant services were as low as a third of those for priced services in the low income countries although near equality in the high income countries). This assumption would reduce the GDP for the lowest income countries by approximately 15 per cent, but would have little effect on the middle and high income countries.

A more realistic measure of the impact of different treatment of the comparison resistant services is to examine the results of the Phase III changes in making comparisons for these services relative to those used in Phase II. The main changes were the use of capital inputs in health and government services, the adjustment for productivity differences in health and education services, and the additions of the number of pupils as a quantity indicator in education. While the revised treatment has an enormous impact on estimates of the quantity of medical services for low income countries, cutting them by 40 or 50 percent relative to the U.S. (KHS 1982, p. 161), the largest effect on GDP was to reduce it by 6 percent (in Kenya and India).

Another insight into the robustness of the estimates is provided by a recalculation of the results of Phases I, III, and IV using as nearly as possible the same methods in all three.[12] The methods were mainly those of Phase IV, the chief exception being that the fixity rule (see below) was not adhered to. The "standardized" results compare with the original benchmark results as shown in Table 1.

The differences produced by the shift to alternative methods are quite small on average; ignoring signs, the mean difference varies from 4 per cent to a little over 6 per cent for the three benchmark surveys. The differences

for individual countries in a few cases are as high as 17 per cent with an outlier at 25 per cent (Malaysia in 1970). Further investigation might reveal the reason for these exceptional cases and make them avoidable.

The net foreign balance. The great preponderance of GDP is absorbed by domestic spending on consumption, capital formation, or government services, but there is often a net positive or negative balance between domestic absorption and production. There is no very clear way to account for the net foreign balance in making comparisons of production or income. These claims are expressed in dollars or other currencies and have no obvious physical counterpart as do other components of final expenditures.

A simple method that has been favored by the Economic Community and used in the more recent phases of the ICP is to convert the net foreign balance to international dollars by use of the exchange rate. This method is not, however, symmetrical with that used for the other categories of final expenditures on GDP. For all of the other categories, an international price (π) is found by relating the PPP for the category to the PPP for GDP as a whole (weighting these ratios by the quantities of the several countries). For each country the quantity in each basic heading is multiplied by the appropriate π. A closer approximation to this method, involving the estimate of a π, was used in the first three Phases of the ICP.[13]

The difference in the estimate of per capita GDP that can be brought about by these two ways of handling the category depends on the size of the π of the net foreign balance, and on the size and sign of the net foreign balance in each country. In the Phase III study, the last in which a price was calculated, each country's net foreign balance converted to U.S. dollars at its exchange rate was multiplied by 1.28, the calculated π. Developing

countries tended to have negative net foreign balances that year (1975), the largest of which was Zambia's, -19.5 percent of its GDP in own currency. If the conversion in the Phase III report to international dollars had been at an I $ price of 1.00 instead of 1.28, Zambia's per capita GDP relative to the U.S. would have been 4 percent higher (i.e., 10.7 percent of the U.S. instead of 10.3 percent). The conclusion is that different treatments of the net foreign balance are likely to have modest effects on comparative real GDP per capita even if the surplus or deficit is very large.

Summary

The main concerns that have been expressed about the benchmark results is that they exaggerate the relative real per capita GDPs of the low income countries. Aside from the errors in the own-currency national accounts, which affect exchange rate conversions equally with PPP conversions, additional uncertainties may arise from improper matching of qualities, sampling variability, and the treatment of problem categories. We do not attempt to measure either the common error of exchange rate and PPP comparisons, stemming from defects in national accounts or, the effects of matching errors. We believe that the former are substantial for some countries and that the major thrust of ICP work to keep the latter small has been successful.

Sampling error cuts both ways and we cannot tell whether the ICP estimate for a given country is too high or too low on this account. We cannot measure the margins of uncertainty arising from the problem categories in any rigorous way. What we can do is to use the sensitivity of the results of the earlier phases to different sources of error and of uncertainty to get some rough approximations of the possible variation.

There are two ways of drawing upon the materials we have presented to obtain some notion of the range of uncertainty in the benchmark results. In the "additive" approach we sum the crude allowances suggested above for sampling variability and for alternative methods. The worst case is that the true estimate will be approximately 10 to 15 per cent lower than the benchmark estimate, the uncertainty consisting mainly of sampling variability. (This refers to those cases in which the sampling error makes the GDP too high.) The uncertainties for comparison resistent services and for the net foreign balance work in offsetting directions to each other.

In the alternative "overall" approach reliance is placed on the difference between the actual benchmark GDP estimates and those that would have been produced by a standardized set of methods that is applied as uniformly as possible to the data of the 1970, 1975 and 1980 Phases. From one side, the worst case for a low income country, leaving the outlier aside, is that the true estimate will be approximately 20 per cent lower than the benchmark estimate in cases in which sampling error is correlated with other sources of differences. In this approach, however, errors or uncertainty arising from methodological factors in the opposite direction are encountered; standardization raises some estimates above the benchmark estimate. The worst case observed in Table 1 for a country in the lowest fourth of the income distribution suggests that the methodological differences could place the true per capita GDP 17 per cent higher than the benchmark estimate. With correlated sampling error, the upside margin could be approximately 25 per cent.

We conclude that margins of uncertainty in the 20 per cent to 25 per cent range, plus and minus, are generous estimates of the outside limits of uncertainty in the benchmark estimates for low income countries originating

from the factors we have examined. Error margins diminish as per capita income rises: for the high income countries (two-thirds or more of U.S. per capita real GDP), they are around 7 per cent, mainly sampling error.

These estimates do not include any allowance for differences owing to the use of different aggregation formulas. Our view is that the Geary-Khamis formula measures what we are seeking to measure from an economic standpoint. If that formula is regarded as merely one of a number of competing formulas, each attempting to answer a somewhat different question, the answer will depend on the formulation of the question.

By contrast with these uncertainties of the PPP conversions, exchange rate conversions seem quite straightforward and free of methodological choices that are difficult to make. One need only, it would appear, take the exchange rate and divide it into the own-currency GDP.

But matters are not so simple. "The" exchange rate that is sought for this purpose is the annual average rate across all transactions, and it is often different from the regularly published "official" rate. When, as is often the case, particularly in developing countries, multiple rates apply to different transactions, or, more commonly, black market rates exist along with fully legal rates, estimates of the average levels of the various rates may be subject to large margins of error, and the relative importance of the official rate and of other rates, necessary to get a weighted average, may involve a large element of guesswork. The resultant uncertainty in the estimate of the average effective exchange rate has, of course, its mirror image in the estimate of real GDP that it is used to derive.

In addition, even the legal rates frustrates erratically, while domestic prices and quantities tend to remain relatively stable. This combination

produces erratic and implausible estimates of real GDP. One way of meeting this problem is to select an exchange rate for a past "equilibrium" year and to extrapolate it to the target year on the basis of relative rates of inflation in the given country and the numeraire country. An alternative is to use a moving average of recent exchange rates. The World Bank, whose World Atlas is the most widely cited source of international income comparisons, uses both of these methods, the former where the official rate seems to deviate by exceptionally large margins. The moving average approach, applied to most countries, presently is based on the exchange rate for the target year averaged with exchange rates for the previous years adjusting the latter rates to the target year by relative rates of inflation.

The differences produced by these and other alternative methods cannot be fully explored here. That they are substantial is indicated by a comparison offered by Ward (1989) of the result of using two different three year periods for averaging. One terminates in the target year, and thus is not centered on the target year; the other centers on the target year but is available only after a further year elapses. When these two methods were applied to 1987 data, three year average rates centered on 1987 ranged from 54 percent to 127 percent of those centered on 1986*; the mean absolute difference for the nine lowest income countries was 20 percent. Thus this is a single source of differences in methodology that produces differences in exchange rate conversions that are larger than those encountered in the PPP converted comparisons.

*This excludes Yugoslavia for which the percentage was 31.

The question of regionalization

In accordance with long held plans, when the need for the development and initial implementation was met and an operational stage reached, the group at the University of Pennsylvania was phased out, and, beginning with Phase IV (reference date 1980), the responsibility for the worldwide comparisons was shifted completely to the U.N. Statistical Office (UNSO). At the same time, outside financial support dwindled and the ability of the UNSO to play a leading role in the comparisons diminished. Also, regional organizations, especially those in Europe, began to produce comparisons for their member countries. As a result, the 1980 comparisons for 60 countries were put together in 7 sets of countries. In Europe, for example, the Economic Community made up one block of 12 countries, and 5 more European countries were compared under the aegis of the Economic Commission for Europe (ECE) with Austria as the center country. Other European countries were added in a set of comparisons prepared by the Organization for Economic Cooperation and Development (OECD). The UNSO took direct responsibility for seven Asian countries for which no regional organization formed a group.

Thus there were sets of regional comparisons, each based on its own average prices, and including some regions within regions (the EC within the OECD). The UNSO linked the regions together through "core country" comparisons. From one to three countries in each group served as core countries, providing prices that overlapped in all groups for each basic heading. For example, France and Spain were core countries for the EC and Kenya and Senegal for Africa. PPPs were estimated for the 20 core countries for each basic heading, using the U.S. as the numeraire country.[14] The other (non-core) countries were linked to the world comparisons through the core country or

countries in their group. This produced a PPP for each basic heading which, with the expenditure data, provided the necessary inputs for a Geary-Khamis calculation for the 60 countries included in Phase IV.

This Geary-Khamis calculation would have provided the final result except for the insistence of Eurostat (the Statistical Office of the European Communities) on the "fixity" principle. Under this rule, the results of the intra-regional comparisons produced by the different regional organizations are not to be altered when the regions are incorporated into the world-wide comparisons. That is, if Germany is 5% higher than France in real per capita income in the EC calculation, the difference must be maintained when the EC countries are put into the context of OECD comparisons or world-wide comparisons. The strong support of the fixity principle by the Eurostat was based on the concern that the production of different relative standings for real GDP per capita for the different pairs of countries would create difficulties for the political and administrative uses of the ICP results.

To implement the fixity principle, UNSO used the results of the 60-country Geary-Khamis calculation to obtain the total GDP for each region. These totals were then distributed among the member countries in each region in proportion to their shares indicated by the within region comparisons. Each country's GDP was distributed to "condensed categories" (akin to the "summary" categories of Phases I-III) and 23 additional aggregations; this distribution was based on each country's distribution as produced by the intra-regional comparison. This method of integrating the results of different regional comparisons (the "GDP consistency" method) has the effect of comparing the quantities of GDP at world prices (I$). However, it has the disadvantage that for a basic heading or condensed category, the sum of the

entries for the countries in the region will not add up to the figure for the region as produced by the world wide comparison. The results are not comparable for countries in different regions for these subdivisions of GDP because they are based on regional rather than world prices. Very large distortions, some in excess of 100%, have been found in relative quantities for basic headings. (Drechsler, 1988) The alternative takes the worldwide quantity calculation for each basic heading or condensed category for each region and distributes it among the countries of the region in accordance with their shares in the results of the worldwide comparison (the "category-control-total" method). This method produces comparability for each subcomponent of GDP (food, etc.), but the components of GDP will not add up to the GDP total estimated directly at world prices.

The effect of the fixity rule has been somewhat mitigated by an agreement allowing international organizations to make available at their discretion the price, quantity and expenditure data for the basic headings. These are not affected by the fixity rule. Thus, users are able to aggregate for themselves the Phase IV and V comparisons at worldwide average prices. Also, the restrictions imposing fixity are to be lifted in about 3 years after publication of the first regional results. Since the Eurostat has put out its results well before the UNSO, the delay for comparisons at worldwide prices imposed by adherence to the fixity rule restrictions may be closer to one year than three.

The fixity principle has serious disadvantages for the worldwide comparisons. It favors within-region comparisons to the detriment of comparisons between countries in different regions. The difficulty is that a different measuring rod (i.e., a different set of relative prices) is applied to

different pairs of countries. Depending on the classification of countries into regions, the German-Japan comparison, for example, might depend on average prices based on EC, European, Asian, and world prices, while Germany/France might be based on European prices alone, and Germany/U.S. on European, North American, and world prices.

This is not to say that regional comparisons should not be made. Their great advantage is that the average prices used to value the products of the member countries are likely to be more characteristic of each country than are world average prices. Regionalization thus diminishes the tendency for the estimates of real GDP per capita to be higher in countries with price structures very different from the one used for the valuation.[15] For some purposes regional or even binary comparisons are appropriate. If, for example, a comparison of the real GDP of the Soviet Union and the United States is desired for strategic purposes, and no other country is concerned in this context, a binary comparison has strong appeal. If on the other hand a comparison of France and Germany is desired for a reason related to the operation of the European Community, a community-wide set of average prices would be more appropriate than either an average French and German prices or average world prices, the latter including the price structures of such diverse countries as the U.S. and India. For general purpose comparisons, however, the use of world average prices seems most appropriate. Given the operating needs of the Economic Community and the general utility of quantity comparisons based on world average prices, it seems sensible to have both regional and worldwide comparisons wherever there is a demand for regional comparisons. It should not be impossible to persuade Economic Community officials and politicians that the Economic Community comparisons are best for

Economic Community purposes even though another set exists.[16]

Extensions to non-benchmark countries

The benchmark comparisons made to date or planned have covered about half of the countries of the world, including all of the populous ones outside of the socialist bloc. The missing countries are almost all LDCs or socialist countries. For many analytical purposes, the covered countries are diverse enough to provide a sample of the distribution of real per capita GDPs in the countries of the world. For some purposes, however, it is important to have income estimates for all of the countries. In this context, benchmark estimates for upwards of 50 countries are missing. While the number of countries covered by the benchmark comparisons may expand, it is unlikely that all the countries will be covered in the near future. Some countries do not have the statistical development needed for participation, and others will be unwilling to devote the resources required. A number of ways have been suggested to prepare estimates of comparative real GDP per capita for these countries by means that entail a smaller expenditure of resources. (See Heston, 1973.) These various approaches to the extension of the estimates to non-benchmark countries are described and illustrated here. It would take a much more ambitious enterprise than can be attempted here to find the optimal application of each approach and to choose among them.

The method closest to the benchmark studies, the "reduced information" approach, calls for carrying out price comparisons on the basis of a much smaller sample of specifications. The full set of the benchmark price comparisons, the rationale goes, contains some that are redundant; that is, their deletion would not alter the result. (Ruggles, 1977) The reduced information estimates would be based on the core of price comparisons that

will yield approximately the same estimate as the benchmark comparisons.

An alternative set of methods is still less costly, since they involve no field work at all. These methods, known as "short-cut comparisons," depend on an estimating equation using the data for the countries that have been included in benchmark studies to form a relationship between real GDP per capita and variables that are widely available for benchmark and nonbenchmark countries. In some versions the independent variables are physical indicators such as steel consumption, and in others they are monetary in character, such as exchange rate converted GDP, or prices used in adjusting cost of living allowances for personnel stationed in different countries. In all of these methods the estimate for each non-benchmark country is obtained by plugging into the estimating equation the values of the independent variables for that country.

Since benchmark estimates are available for virtually all of the developed countries, this exploration of the various methods has been confined to predicting benchmark estimates for developing countries.

Equations based on benchmark countries that were developing countries did not yield predictions of benchmark results superior to those derived from equations based on all the countries, and only the latter are reported here. The equations were used to estimate domestic absorption, and GDP was then obtained by adding the net foreign balance (converted to dollars via the exchange rate). (See Summers and Heston, 1984.)

Reduced information estimates

The reduced information method, although considerably cheaper than the full benchmark survey method still involves substantial cost and thus has been infrequently attempted. A major effort in this area was an experimental study

of 13 developing countries by Sultan Ahmad of the World Bank. Ahmad (1988) began by experimenting with ICP Phase III data to find out the minimum number of price comparisons which could satisfactorily explain the observed values of real GDP per capita in the benchmark countries. He identified a set of such price comparisons for 126 individual products in about 30 categories of GDP expenditures. Ahmad's estimates for the countries in his experiment range from 68 per cent of the benchmark result to 103 percent. The mean absolute deviation is around 10 percent. There is a tendency for the deviation from the benchmark results to be smaller for the higher income countries. In every case the reduced information estimate is closer to the ICP benchmark than is the exchange rate converted GDP. (See Table 2.) Ahmad finds his reduced information also comes closer to matching the benchmark estimates than a short cut set of estimates produced by Summers and Heston using nominal GDP per capita and openness as independent variables.

The experiment seems to support further exploration of reduced information methods. It should be investigated whether they work well for particularly small or especially poor countries for which benchmark studies cannot be readily carried out, or for which the costs of benchmark studies would be too high.

Short cut methods: physical indicators

There is no very strong theory underlying the physical indicator approach, although lurking in the background is the idea of Engel curves. That is, for most goods, consumption is correlated with income. Here, the estimate of the level of income is made to depend upon the level of consumption, but that is a small matter since the real purpose of the exercise is not to explain the economic world, but simply to predict the real level of GDP.

In the physical indicator approach, the relationship between real GDP per capita and each of perhaps a score or more of indicators is examined. (See, for example, Beckerman, 1966 and U.N. Economic Commission for Europe, 1980.) One tactic is to screen the indicators to identify those with high simple correlations with real GDP per capita. Then alternative combinations of three or four of these physical indicators with high simple correlations are correlated with real GDP per capita to find the combination which yields the highest $\bar{R}^2$. (The multicolinearity among the indicators is so high that no more than a few variables add to the degree of explanation.) An alternative tactic is to feed all the indicators for the set of benchmark countries into the computer and to allow the computer to perform a stepwise regression specifying a cutoff when added variables no longer reduce the unexplained variance by a stipulated amount. A disadvantage of the stepwise regression method is that the results are influenced by the order in which the variables are introduced into the regression.[17]

Short cut methods: monetary indicators

Another set of short cut methods uses monetary indicators; that is, those relating to nominal or exchange-rate converted GDP. Additional variables x_j include openness, price isolation, education, money growth, mineral share and the trade balance (Kravis, Heston, and Summers, 1978; Kravis and Lipsey, 1983; Clague 1986).

The systematic relationship found to exist between real PPP converted GDP (r) and nominal (exchange rate converted) GDP (n) provides the rationale for the monetary indicator approach. Briefly, prices are high in rich countries because services are relatively expensive. That is either because the rich countries' margin of superiority in productivity is lower in services

than in goods, or because labor is expensive in rich countries and services are relatively labor-intensive, making services expensive. (Kravis and Lipsey, 1983; Bhagwati, 1984.) This circumstance together with the fact that traded goods prices tend to be closer to uniformity in different countries creates higher price levels for GDP in rich countries. The consequence is that the relationship of r to n falls as n increases. Once again, although the line of causation presumably runs from r to n, r is treated as a function of n for purposes of extrapolation.

Short cut methods: price indicators

The integration of the world economy and the rise of international government and of multinational business enterprises has led to the existence of international organizations that have employees stationed in different countries. In view of wide differences in living costs, the organizations often maintain a system of "post-adjustment" allowances to equalize the real incomes of personnel of equal status in the headquarters locations and those stationed elsewhere. For some organizations with personnel in many locations this work involves a rather elaborate system of price comparisons. The U.S. Department of State makes price comparisons for about 150 cities (U.S. Department of State, 1981) and the United Nations for about 125 cities (U.N., September 1980). Private organizations also produce comparisons for a large number of locations. These price comparisons may be treated as proxy PPPs in an estimating equation for the real domestic absorption of benchmark countries.[18] Estimates for non-benchmark countries may then be derived by inserting their comparative prices as shown in the post adjustment data. The assumption underlying this approach is that the difference between the national price level and the price level encountered by foreign personnel

dwelling in the capitol or other leading city for professional or business reasons, is the same everywhere. Some experiments with these different sources showed that they produced very similar results when they were used as PPPs to convert GDPs in own currencies to real GDP per capita. For brevity and simplicity, we report here only on the results using the U.N. data.

A comparison of the results of the shortcut experiments

In Table 3, the results of our brief essay into these shortcut methods are presented; the shortcut results are expressed as ratios of the benchmark estimates. For comparative purposes, the exchange rate conversions are also presented.

The basic procedure was to array the developing countries in the 1980 (Phase IV) in order of PPP converted per capita GDP, and to use each half of the sample (odd ranks and even ranks) to predict a 1980 estimate for the countries in the other half. In a very preliminary comparison of methods, the price indicator approach appears to produce marginally better results than the monetary approach, and both are better than the physical indicators. The price indicator method has the advantage over the monetary approach in that, unlike the latter, it is not dependent on exchange rate and related policies. The predictions are on the average within 15% of the benchmark estimates, but the range is from predictions 46% above the benchmark to predictions more than 30% below. However, the other shortcut methods produced average errors of 20 per cent or more and wider ranges of error.

All the shortcut estimates clearly outperform the exchange rate conversions; the latter are characterized by mean deviations of 35% and understatements of relative GDPs by as much as 70 per cent or more.

Extensions to other years

The emerging pattern in which benchmark studies are made in quinquennial years leaves open the question of estimates for the in-between years. Also not every benchmark country participates in every benchmark year and the need for extrapolating the available benchmark estimates to other years, benchmark and nonbenchmark, arises in these cases too.

A rough and ready extrapolation of the benchmark year estimate for real GDP per capita relative to that of the numeraire country is possible at the aggregate level of GDP per capita. The benchmark year real GDP per capita for the country and the numeraire country are simply multiplied by the ratio of extrapolation year to benchmark year real per capita GDP from national data. An alternative procedure is to extrapolate the PPP of the given country and divide the result into the extrapolation year current price GDP per capita; the resulting estimate in dollars may be put in index number form by dividing by the per capita GDP of the U.S. in current (extrapolation year) dollars.

The disadvantage of these simple procedures is that the growth rate in the given country's GDP between the benchmark year and the extrapolation year is measured using the relative prices of that country, whereas the growth rate of the numeraire country GDP is measured using its (different) relative prices.

At the other extreme, one can envision extrapolating each price from the benchmark year to the extrapolation year, recalculating the PPPs for the basic headings, and using these PPPs in conjunction with the extrapolation year expenditure breakdown, to produce a Geary-Khamis calculation This would be very close to producing a new benchmark study. If the PPP for each basic heading rather than each price were extrapolated, differences among countries'

price structures would still have an impact but it would be confined to the influences within the basic headings; for combining the basic headings, the average price structure would be used.

The extrapolations can be made to obtain the comparisons in either current year or in benchmark year international prices. In the former case, the PPPs are extrapolated and divided into the current expenditure values in order to obtain quantities in current prices. These quantities expressed in international dollars can be summed to obtain aggregates up to total GDP per capita in current year international prices. If on the other hand the aim is to estimate extrapolation year quantities in benchmark year international prices, the quantities are extrapolated directly and the values for the individual categories added to obtain whatever aggregation is desired.

A question that remains to be investigated is the time span over which extrapolations can be made without introducing very substantial differences from benchmark estimates. The implicit assumption in the scheme for quinquennial estimates is that a five year period is not too long. As experience in benchmark comparisons accumulates it will be possible to determine whether this or a longer interval is feasible.

An alternative being followed by the EC is to have annual estimates of a benchmark character achieved by doing the benchmark comparisons segment by segment over a three or five year cycle. Benchmark type comparisons might be made for food in one year, other consumer nondurables the next year, etc. Extrapolated values would be filled in for those components not covered in the pricing of that year. This has the merit of affording an opportunity to integrate ICP data with the price and quantities indexes and national accounts work of the participating countries. It is relevant to observe here that the

work on the ICP by many developing countries has proved to be a valuable aid in strengthening their domestic statistical systems. The disadvantage of complete comparisons at five year or longer intervals is that institutional memories are short. Studies made five years apart may require a great deal of new learning all over again. Continuity might help domestic statistical work as well as international comparisons.

A set of international comparisons of real per capita products and of price levels for 130 countries annually from 1950 to 1985 has been offered by Robert Summers and Alan Heston (1988). Using the breakdown of total GDP per capita into consumption, government, capital formation and the net foreign balance, they extrapolate benchmark year comparisons backwards and forwards to other years in order to derive estimates in current prices of each year as well as in 1980 international prices. They show breakdowns for consumption, investment, and government both in current prices and 1980 prices and also the price levels for each of these components. (Price level is the PPP divided by the exchange rate.) The data for 1980 are printed out in a table and data for all years are made available in the form of diskettes.

A problem of consistency arises for the Summers-Heston effort for countries which there have been two or more benchmark comparisons.

Consistency requires that: $\frac{y_{t+5}}{y_t \cdot g} = 1$

where y_t is the per capita GDP (in current international prices) of a given country relative to the numeraire country in the year t, say 1975, and the year y_{t+5} is the same for a benchmark five years later (1980), and g is the relative growth rate of the given country relative to that of the numeraire

country between t and t + 5, as obtained from national accounts data.

Summers and Heston achieved consistency by an ingenious method that first measures the deviation from consistency and then decomposes it into the amounts attributable to y_t, $y_t + 5$, and g. On this basis an adjustment factor is derived for each of the three elements.

The reconciliations alter some of the benchmark estimates to a notable degree; the largest decline in a 1980 benchmark is about 8%, and the largest increase about 9%. On average, however, the ups and downs virtually cancel out.

A limitation of these calculations -- unavoidable without a great deal of work with each country's national accounts often requiring data not in the public domain -- is that g is based on a mixture of domestic and international prices. The reason is that the extrapolation of y_t in 1980 has to be done for practical reasons (the unavailability of price indexes) not for 150 benchmark components, or even for 35 summary categories but for a breakdown of GDP into only four major subdivisions. For consumption, for example, the extrapolation is based on each country's own GDP consumption deflator, embodying the country's own prices and weights for the goods and services that make up its consumption total. More generally, each of the four major subdivisions receives an international price relative to the other three. But the extrapolation necessary to prepare the inputs of PPPs of the Geary-Khamis aggregation is carried out with a purely domestic index. Furthermore, not only is g affected by the intermingling of domestic and international prices, but so are the adjustments made to the benchmark estimates. (It should be added that there are complications in this work that we will not try to deal with, but with which Summers and Heston cope in careful and plausible ways.) One possible objection, easily remedied, is that estimates for the later benchmark

years are extrapolated backward to measure the inconsistency from the earlier benchmark year and the earlier year estimates are extrapolated forward to obtain a second equally meritorious measure of the inconsistency, and an average of the two is taken. This introduces ambiguities about the date of the price structure which is being applied.

Even if this were to be corrected, the massaging of the benchmark comparisons and of the growth rates between them to make them consistent with one another further diminishes the transparency of the prices that are used to evaluate the quantities. The price structures of the benchmark year and of the domestic deflators -- representing different things -- are meshed together.

For some users, the advantage of having consistency between time to time and place to place data will compensate for the ambiguity in what is being measured. For others, growth rates based on international prices may be so attractive, that they are willing to overlook the limited role that international prices actually play in the calculations.

However our view is that the best general purpose estimates of growth rates are those derived directly from the domestic price deflators of the countries. They have relatively clear conceptual underpinning. (They are, to be sure, made less comparable from country to country by the use of different base years.) Similarly we think that the best estimates of real GDP per capita levels are those produced by the benchmark studies, unaltered by modifications based on a mixture of domestic and international prices.

Having said this, we add that the Summers and Heston's is the most comprehensive set of PPP-based estimates that exists; their "consistentized" data cover almost all countries and 35 years. As noted, their data are aimed at uses requiring consistent estimates of levels and of changes in output.[19]

Growth rates

The usual way of calculating growth rates is to take the changes in the real GDP of each country as measured by its own implicit deflator. Thus, as noted in earlier discussion of extrapolation to other years, the growth rate of the given country measures the change in a basket of goods that is different from that measured by the growth rate of the numeraire country; also, different (price) weights are assigned to overlapping goods in the two countries. Such comparisons answer the question "How much change has there been over the period in the quantity of the base (current) bundle of goods produced in country one compared to the change in the quantity of the different base (current) bundle produced in country two?" Such growth rates have the merit of dealing with a basket of goods that reflects the preferences of purchasers of final product in one of the years being compared. (In the language of the ICP, they have the desirable property of characteristicity.) They have the disadvantage (?) that an equal growth in two countries in the quantity of a given good may be counted as contributing more to aggregate growth in one country than in another.

Comparisons of growth rates based on international prices of a given year answer the question "How much change has there been over the period of the total quantity of goods absorbed in country one, compared to the change in the total quantity of goods produced in country two, recognizing that the list of goods may be different in each situation, but valuing the goods as the same set of world average prices?" Such growth rates have the merit of treating a given increase in a given good as making the same contribution to growth in both countries. They have the drawback that the prices used may be very dissimilar from the prices of one or both of the situations.

The choice between the two approaches depends on the use to be made of the growth rates. If the purposes are closely related to welfare considerations, own price growth rates are preferable because they are more closely related to the choices confronting the purchasers of final product in each country. If, on the other hand, the purposes are related to production, it may be argued that international price growth rates are preferable. It can be claimed that the international average prices are more closely related to world opportunity costs; each country has the opportunity to acquire a given good from the rest of the world at its international average price[20].

Future ICP work

Phase VI (reference date, 1990) is encountering difficulties because a number of the binary comparisons which were intended to be used to link regions appear to be languishing. There is a great danger that Phase VI will be a series of regional comparisons not completely without links but without enough links to produce systematic worldwide comparisons. A way should be found to revive the prospects for worldwide comparisons that does not require large resources.

One possibility is to modify the comparison strategy from a core country approach to a core product approach. All countries in a region would be encouraged to include some specifications that are easy to price but which are found also in other regions. At the same time, one or two countries in the region would try to do more extensive matching with a "reference" country outside the region.[21]

As already suggested, a consequence of the reliance on regions to organize much of the data in Phases IV and V has been the emergence of some difference in methods, although the broad strategy of the ICP and most of its

detailed methods were adhered to. A notable difference was a large quality adjustment made in the central European comparisons for physicians; the productivity of each physician in Hungary, Poland, and Yugoslavia, was taken as 50% of that of the Austrian physician. This seems difficult to reconcile with the other European comparisons, in which no adjustment was made for physicians. The question arises whether the relationship between the productivity of physicians in say Turkey and Germany, included in the OECD comparisons without adjustment, was consistent with treatment of the productivity of Hungarian and Austrian physicians in the Austrian-centered comparisons.

The differences are greater than they would have been had the UNSO the resources to coordinate the work of the different regions, and the differences will grow larger unless the UNSO takes a more extensive role. UN experts on consumer goods pricing, capital goods and construction should attend at least one of the planning meetings of every regional group. The UNSO experts could encourage the regions to include some specifications that can be matched in other regions and to provide the UNSO with data that enable it to do the worldwide comparisons by a standard set of methods. Continuing efforts should be made to reach a world consensus on the problem points, but as long as the regions supply the UNSO with the standard data set, they should be free to use methods in their own regional comparisons that differ.

Another important role that a UN presence at planning meetings might fulfill is to discourage any tendency to influence the results by manipulating the inputs. An objective outside presence in the course of data collection and processing might further diminish this possibility.

In addition to the benchmark work, there is a need for more research into methods of extending the benchmark estimates to non-benchmark countries.

Conclusion

A system of international comparisons of income and of the purchasing power of currencies covering most countries and currencies for the period 1950-85 is now in place. The estimates, which are on an annual basis and include breakdowns for consumption, capital formation and government, rest on a relatively small number of benchmark comparisons. The latter include nearly 80 countries, some for single years, the earliest of which is 1970, and others for as many as five years the latest of which is 1985. The benchmark estimates are based primarily on price comparisons which have been produced by a world-wide cooperative effort involving many countries and international governmental agencies.

The income comparisons relate to GDP as defined in the U.N. System of National Accounts. They are derived by applying a set of world average prices to the quantities comprising each country's national absorption of final goods (and net claims against foreigners). If this approach is accepted, the benchmark results are not very sensitive to plausible alternative treatments of certain methodological issues, to the resolution of which theory gives little guidance. (The guide to "plausible" alternatives includes those seriously considered or adopted by the UNSO, Eurostat, ECE, or the OECD.)

We conclude that margins of uncertainty in the 20 per cent to 25 per cent range, plus and minus, are generous estimates of the outside limits of uncertainty in the benchmark estimates for low income countries originating from the factors we have examined. The margins narrow as per capita income rises; for high income countries, they may be around 7 per cent.

If the Geary Khamis formula is not regarded as uniquely suitable for the comparisons, the results will be further affected by the formulation of the

question to be answered and the appropriate aggregation formula.

Larger margins of uncertainty attach to estimates for countries that have not been covered by benchmark studies. The worst case possibilities may be as high as 35% on the upside and 30% on the downside. These are only rough approximations. They are probably susceptible of narrowing with further research.

Even were the benchmark and nonbenchmark estimates of uncertainty margins to prove too small, they have to be weighed against the errors involved in the use of exchange rate conversions, the only alternative to conversions via PPPs. For the very poorest countries the PPP conversions yield estimates of per capita GDP that are more than three times the exchange rate converted figure, and for the developing countries as a group the average ratio is over two times. Exchange rates can only be justified as converters if they reflect the relative purchasing power of currencies better than the ICP PPPs do. The current literature on exchange rate determination with its stress on capital movements and expectations, and the recent volatility of exchange rates accompanied by relatively sluggish movements of domestic prices, make clear the inadequacy of exchange rates as PPPs. The exchange rate converted figures for the poorest countries are far from the lowest PPP-converted estimates of real GDP per capita that emerge from the consideration of the uncertainties inherent in that method.

Furthermore, the results of exchange rate conversions are even more sensitive to the methods employed than is the case for PPP conversions. The notion that exchange comparisons rest on a simple and transparent procedure using standard data produced by the operation of markets is illusory.

Uncertainties are inherent in international comparisons. Those based on PPP conversions can be reduced by the investment of further resources, especially by extending benchmark estimates to more countries (particularly low income ones) or years, and time may bring further consensus on methodological questions and narrow the uncertainty range.

The future of ICP estimates appears assured in Europe, particularly in the European Communities. In other regions, prospects vary, but at present the outlook for systematic worldwide comparisons for Phase VI (1990) do not look bright. It will take a renewed impetus which in the circumstances can only be provided by the UNSO and the World Bank, to establish an ICP with comprehensive coverage on an ongoing basis.

There has been an international effort, stretching over the better part of the half century to develop comparability in the national accounts of the various countries (i.e., the SNA). It would be ironic to lose momentum gained with great effort toward the final step in establishing comparability -- the translation of own-currency GDPs into comparable measures of real income.

1. PPP is defined here as the member of units of currency j required to purchase the same amount of goods as a unit of the numeraire currency can purchase.

2. See Kravis, Heston, and Summers, 1982, for a fuller account. The methods have been summarized in a number of papers (see Kravis, 1984 for example). The present outline is provided to enable readers not familiar with the methods to understand the discussions that follow.

3. In a few categories, quantities are compared and the price comparisons are derived by dividing the quantity ratios into the expenditure ratios. Also, comparisons in some regions have been based on a more detailed breakdown and others on a less detailed breakdown. The basic headings represent the most detailed breakdown of expenditures that it is possible to make for many countries. An effort has been also made to define them so that they include products that are alike with respect to price determining influences.

An approach which is more difficult to implement is to build up the GDP comparison in terms of the industries producing the output. See Paige and Bombach (1959) and Maddison and van Ark (1987).

4. The Country-Product Dummy method (CPD) does this through a regression in which the log price of an item is the dependent variable and the independent variables are two sets of dummy variables, one for the various countries and the other for the different specifications. The coefficient of the dummy variable for a given country represents the log of the PPP for the category in that country relative to the numeraire country.

See Summers, 1973 and Kravis, Heston and Summers, 1982. The other method of coping with differing price lists, offered independently by Elteto,

Koves, and Szulc (hence, the "EKS" method), has been used in the European Comparisons. The EKS index for a given pair of countries is the geometric mean of the direct Fisher indexes (weighted twice), and all the bridge country Fisher indexes:

$$EKS_{jk} = \left[F_{jk}^2 \cdot \prod_{\substack{i=1 \\ i \neq j,k}}^{n} \frac{F_{jl}}{F_{kl}} \right]^{1/n}$$

where F is the Fisher Index: j,k, and l are countries; and n is the number of countries. (A bridge country, l in this case, is one that links together two countries through a comparison of each with the bridge country. For example, a bridge country method for obtaining F_{jk} is to divide F_{jl} by F_{ki}.) The CPD and EKS methods generally produce similar results. Krijnse-Locker, 1982.

5. The base country serves merely as a numeraire. It makes no difference for the quantitative relationship among the countries which one serves as the numeraire.

6. For example, $I_{j/k} = I_{j/l} \div I_{k/l}$ where I is an index of quantities or prices and j, k, and l are countries. This ensures that the relative positions of the countries will be unambiguous.

7. In a table with countries in the columns and categories in the rows, the entries show the correct relative quantities on any row and are additive in the columns to any desired aggregate such as consumption or GDP. This table is akin to the familiar national accounts time to time table showing final expenditures in constant prices.

8. Characteristicity is indeed promoted by the fact that half the weights in a Fisher index refer to the given country's own prices or quantities. However, the characteristicity on this account may be offset in the cases of partners with price or quantity structures that differ radically from those of the given country.

9. See Statistical Office of the European Communities; Organization for Economic Cooperation and Development; U.N. Statistical Office. 1989.

10. These calculations are based mainly on the 34 Phase III countries. An increase in the number of countries studied might turn up some larger differences.

11. Based on 34 countries (KHS, 1982, p. 194) In all of the Phases, services in health and education provided at public expense are included in "consumption" rather than "government" so as to make the country to country comparisons of these aggregates invariant to the source of their financing.

12. Unpublished data kindly made available to us by Professors Heston and Summers.

13. For each country, the ratio of the exchange rate to its PPP as estimated from a preliminary Geary-Khamis calculation for GDP, excluding the net foreign balance, was used to form the international price. In this calculation the ratios for the different countries were weighted by the relative importance of each country in total GDP. The method of Phases IV and V, in which the net foreign balance is not multiplied by an international price, is equivalent to taking that price as equal to one.

14. The CPD method was used.

15. The theoretical case for regionalization rests on grouping countries with similar price and quantity structures together. Geographical propinquity, while an obvious starting point in grouping countries, is not an adequate criterion. In the real world, the regionalization that is demanded is heavily influenced by political considerations, and some "regional" groups cut across continental lines (OECD) and others subdivide continents (EC). (See KHS, 1982.)

16. In fact, the fixity rule makes little difference in the relative per capita GDPs of the high income countries but can have notable effects on a few middle and low income countries. This statement is based on a comparison of the results for the 60 countries of Phase IV (1980) published by the UNSO and Eurostat which embody fixity, and the estimate of GDP obtained by summing the international dollar values of the 151 basic headings available on a UNSO tape. None of the differences for the high income countries exceeded 4.5 percent, while the difference was 5 to 10 percent for 12 other countries and over 10% for 8 others (with the largest difference 21 percent).

17. A further difficulty is that many indicators are not available for all of the countries in the benchmark sample. Indicators that have a large number of missing observations are dropped. For the remaining cases, it is possible to run separate regressions for the benchmark countries in which all of the indicators are present and for those in which all but one, all but two, or all but three are present. For each non-benchmark country, per capita GDP is estimated from the equations with the largest number of independent variables available for that country. This procedure would mean that the estimating equation used for real GDP per capita varies from one country to another.

18. Robert Summers has suggested that this might be regarded as a simplified reduced information approach.

19. For some users the presentation of the original benchmark data in their tables would also have been valuable. The omission of own-country growth rates is not a disadvantage because they can be readily computed from widely available summaries of the national accounts data of different countries -- e.g., IMF, International Financial Statistics.

20. For an analysis based on a different concept of world prices see Bhagwati and Hansen.

21. CPDs would be calculated for each region and used to fill in missing prices in the regional set. Then a second-stage CPD would be calculated, covering all included countries, each with a complete set of the prices used in its region. The PPPs derived from the second-stage CPD would be the inputs for a Geary-Khamis calculated for all the countries. Thus advantage would be taken of the similarity of price structures within regions to cope with the missing price problem, but the desired properties of multilateral comparisons would be retained. A single set of international prices would be used to value each country's quantities (income).

Table 1

Standardized Benchmark Estimates of Real GDP Per Capita

Compared to Original Benchmark Estimates

	Ratios, Standardized/Original					
	1970		1975		1980	
	All Countries	Lowest Income Fourth	All Countries	Lowest Income Fourth	All Countries	Lowest Income Fourth
No. of countries	16	4	34	8	60	15
Range:						
Minimum	.75	.89	.87	.91	.81	.89
Maximum	1.04	.96	1.04	1.03	1.17	1.14
Mean Absolute Deviation from 1.0	.059	.061	.024	.031	.057	.061

Table 2

Estimates of Per Capita GDP by Reduced Information Method Compared with Benchmark Estimates, 1980

(US=100)

	Exchange Rate Converted (1)	Reduced Information (2)	Bench-mark (3)	Ratio to Benchmark: Exchange Rate Converted (1) ÷ (3) (4)	Ratio to Benchmark: Reduced Information (2) ÷ (3) (5)
Costa Rica	18.5	21.3	27.2	.68	.78
Dom. Rep.	10.4	15.2	16.8	.62	.90
Guatemala	9.5	14.4	17.6	.54	.82
India	2.1	4.9	5.3	.40	.92
Indonesia	4.3	8.7	9.6	.45	.91
Kenya	3.7	5.4	6.0	.62	.90
Morocco	7.7	11.0	10.7	.72	1.03
Nigeria	8.7	6.9	7.5	1.16	.92
Panama	15.9	23.5	23.9	.67	.98
Senegal	4.6	4.9	7.2	.64	.68
Tanzania	2.2	3.0	3.2	.69	.94
Tunisia	11.9	16.9	16.9	.70	1.00
Zimbabwe	6.3	7.2	7.8	.81	.92
U.S.A	100.0	100.0	100.0	1.00	1.00

Source: Ahmad, 1989

Table 3

Predictions of LDC's 1980 Real GDP per Capita by Various Short Cut Methods

Set of countries upon which regressions are based	Estimates Produced for Developing Countries in the Opposite Set as Ratios to Benchmark Estimates		
	Range		Mean Absolute
	Maximum	Minimum	Deviation
	I. Physical Indicators		
Odd Numbered	1.811	.313	.213
Even Numbered	1.771	.737	.183
	II. Price Indicators		
Odd Numbered	1.335	.714	.145
Even Numbered	1.459	.682	.102
	III. Monetary Indicators		
Odd Numbered	1.516	.698	.164
Even Numbered	1.574	.571	.142
	IV. Exchange Rate Converted		
Odd Numbered	1.291	.279	.294
Even Numbered	1.323	.223	.318

Note:

The Phase IV countries were arrayed according to increasing real GDP per capita as measured in the benchmark study, and divided into two sets (odd and even numbered countries). For the countries in each set, a regression equation was estimated with per capita domestic absorption as the dependent variable and the various indicators in I, II, and III as the independent

Table 3 (continued)

variables. (In some equations data were not available for all countries.) Each equation was then used to "predict" the 1980 real GDP per capita (adding the net foreign balance) for the developing countries in the opposite set. The "predictions" are compared with the benchmark estimates in the three columns of the table. The figures in IV are exchange rate converted per capita GDPs.

Appendix Table
Equations Used to Generate Predictions Summarized in Table 3

I.) Physical Indicators

A. $DA_{pc} = \underset{(0.36)}{.763} + \underset{(1.56)}{.193}$ (Calorie Intake p.c.) + $\underset{(2.82)}{1.698}$ (Life Expectancy)

+ $\underset{(2.95)}{.225}$ (Energy Cons. p.c.) - $\underset{(3.90)}{.429}$ (% of Labor Force in Agriculture)

$\bar{R}^2 = .95$

B. $DA_{pc} = \underset{(1.42)}{-3.899} + \underset{(1.74)}{.543}$ (Calorie Intake p.c.) + $\underset{(2.43)}{.201}$ (Secondary School)

+ $\underset{(3.21)}{1.501}$ (Life Expectancy) + $\underset{(3.67)}{.194}$ (Energy Cons. p.c.)

- $\underset{(4.17)}{.237}$ (% of Labor Force in Agriculture) $\bar{R}^2 = .97$

II) Price Indicators

A. $DA_{pc} = \underset{(8.46)}{2.18} + \underset{(23.16)}{.744}\ (DA_{UN}) - \underset{(2.44)}{.221}\ (D_A)$ $\bar{R}^2 = .97$

B. $DA_{pc} = \underset{(11.25)}{2.448} + \underset{(27.02)}{.742}\ (DA_{UN}) - \underset{(5.44)}{.397}\ (D_A)$ $\bar{R}^2 = .98$

III) Monetary Indicators

A. $DA_{pc} = \underset{(9.79)}{1.305} + \underset{(21.17)}{.691}$ (n) - $\underset{(0.54)}{.036}$ (Openness) - $\underset{(4.64)}{.442}\ (D_A)$ $\bar{R}^2 = .97$

B. $DA_{pc} = \underset{(12.75)}{1.641} + \underset{(19.61)}{.620}$ (n) + $\underset{(1.61)}{.129}$ (Openness) - $\underset{(6.93)}{.649}\ (D_A)$ $\bar{R}^2 = .97$

Appendix Table (continued)

Notes:

The predictions for developing country real GDP per capita that are generated by these equations are summarized in Table 3.

The regressions are based on all the Phase IV countries, and are in log form.

t-statistics in parentheses

A = equations for odd numbered countries

B = equations for even numbered countries

The variables are on a per capita basis where appropriate (e.g., energy consumption, domestic absorption, etc.)

DA_{pc} = domestic absorption in international prices, from benchmark studies

DA_{UN} = own currency domestic absorption ÷ PPP from UN post adjustment data

D_A = dummy variable with value of 1 for African countries; 0 for others.

n = exchange rate converted GDP per capita

OP = exports plus imports/GDP

The prediction of real GDP per capita was obtained for each country by adding the net foreign balance to the estimate of DA_{pc} produced by the equation.

References

Ahmad, Sultan. (1988) "International Real Income Comparisons with Reduced Information," J. Salazar-Carrillo and D.S. Prasada Rao, Editors, World Comparisons of Incomes, Prices, and Product, Elsevier Science Publishers, B.V. (North Holland).

Ahmad, Sultan. (1980) Approaches to Purchasing Power Parity and Real Product Comparisons Using Short-Cuts and Reduced Information, World Bank Staff Working Paper, September, #418.

Beckerman, Wilfred. (1966) International Comparisons of Real Incomes, Development Center, Organization for Economic Cooperation and Development, Paris.

Bhagwati, Jagdish N. (1984) "Why Are Services Cheaper in Poor Countries?," Economic Journal, June, Vol. 94, No. 374, pp. 279-286.

_____, and Hansen, Bent. (1973) "Should Growth Rates Be Evaluated At International Prices?" J. Bhagwati and Richard S. Eckaus, Editors, Development and Planning.

Clague, Christopher. (1986) "Short Cut Estimates of Real Income," Review of Income and Wealth, September, pp. 313-331.

Drechsler, Laszlo. (1988) "The Regionalization of and Other Recent Developments in the U.N. International Comparison Project," J. Salazar-Carillo and D.S. Prasada Rao, Editors, World Comparisons of Incomes, Prices, and Product, Elsevier Science Publishers, B.V. (North Holland).

Gilbert, Milton and Kravis, Irving B. (1954) An International Comparison of National Products and the Purchasing Power of Currencies: A Study of the United States, the United Kingdom, France, Germany, and Italy, Paris: OEEC.

Heston, Alan W. (1973) "A Comparison of Some Short-Cut Methods of Estimating Real Product Per Capita," Review of Income and Wealth, March, 19[1], pp. 79-104.

Hill, T. Peter. (1977) "On Goods and Services," Review of Income and Wealth, December, 23[4], pp. 315-38.

_____. (1981) "Multilateral Measurements of Purchasing Power and Real GDP," Report prepared for the European Communities Statistical Office and the Economic Commission for Europe. Luxembourg.

Intersecretariat Working Group on the International Comparison Programme (ICP). (1989) "Report of the Meeting of the Intersecretariat Working Group on the International Comparison Programme (ICP)," July, New York.

Isenman, Paul. (1980) "Inter-County Comparisons of 'Real' (PPP) Incomes: Revised Estimates and Unresolved Questions," World Development, January, 8[1].

Kravis, Irving B. (1984) "Comparative Studies of National Incomes and Prices, "Journal of Economic Literature, March, Vol. XXII.

Kravis, Irving B., Heston, Alan W. and Summers, Robert. (1978a) International Comparisons of Real Product and Purchasing Power, Baltimore, MD: Johns Hopkins University Press.

_____. (1978b) "Real GDP Per Capita for More Than One Hundred Countries," Economic Journal, 88(350), pp. 215-42.

_____. (1982) World Product and Income: International Comparison of Real Gross Domestic Product, Baltimore, MD: Johns Hopkins University Press.

_____; Kennessey, Zoltan; Heston, Alan W. and Summers, Robert. (1975) A system of International Comparisons of Gross Product and Purchasing Power, Baltimore, MD.: Johns Hopkins University Press.

_____; Lipsey, Robert E. (1983) "Toward an Explanation of National Price Levels," Princeton Studies in International Finance, November, No. 52.

Krijnse-Locker, Hugo. (1982) "On the Estimation of Purchasing Power Parities on the Basic Heading Level." Paper presented at the International Association for Research in Income and Wealth, Special Conference on Purchasing Power Parities, Luxembourg, September 21-24.

Maddison, Angus and Bart van Ark. (1987) The International Comparison of Real Output Purchasing Power and Labor Productivity in Manufacturing Industries: A Pilot Study for Brazil, Mexico and the USA for 1975, Institute for Economic Research, Faculty of Economics, University of Groningen.

Paige, D. and G. Bombach. (1959) A Comparison of National Output and Productivity of the United Kingdom and the United States, Organization for European Economic Cooperation, Paris.

Ruggles, Richard. (1977) The Wholesale Price Index -- Review and Evaluation, Council on Wage and Price Stability, Council report, Washington, D.C., June.

Statistical Office of the European Community. (1977) Comparisons in Real Values of the Aggregates of ESA, 1975. Luxembourg: EEC, 1977.

_____; Organization for Economic Cooperation and Development; United Nations Statistical Office; Expert Group on ICP Methodology (June, 14-16 1989) Paris; Report of the Expert Group on ICP Methodology, (June, 6-10 1988) Luxembourg.