Volume Title: International Economic Transactions: Issues in Measurement and Empirical Research

Volume Author/Editor: Peter Hooper and J. David Richardson, editors

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-35135-1

Volume URL: http://www.nber.org/books/hoop91-1

Conference Date: November 3-4, 1989

Publication Date: January 1991

Chapter Title: The International Comparison Program: Current Status and Problems

Chapter Author: Irving B. Kravis, Robert E. Lipsey

Chapter URL: http://www.nber.org/chapters/c8437

Chapter pages in book: (p. 437 - 468)

12 The International Comparison Program
Current Status and Problems

Irving B. Kravis and Robert E. Lipsey

12.1 Introduction

A worldwide United Nations program to produce international comparisons of real GDP and its components and the purchasing power parities (PPPs) of currencies has been under way for twenty years.[1] 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 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.

Irving B. Kravis is university professor emeritus at the University of Pennsylvania and a research associate of the National Bureau of Economic Research. Robert E. Lipsey is professor of economics at Queens College and the Graduate Center, City University of New York, and a research associate of the National Bureau of Economic Research.

The authors are grateful for helpful comments by Alan V. Deardorff and Alan Heston, neither of whom is responsible for the views expressed here. The statistical work for this paper was done by David Robinson, and the manuscript was prepared by Maryellen Sykes.

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

12.2 An Outline of ICP Methods

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).[2] GDP is intended in concept 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 incomplete 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,

$$\frac{Q_j}{Q_b} = \frac{E_j}{E_b} \div \frac{\text{PPP}_j}{\text{PPP}_j}$$

where j and b are countries, Qs are physical quantities and Es are expenditures in own currency.

2. See Kravis, Heston, and Summers (1982) for a fuller account of ICP methods. The methods have been summarized in a number of papers (see, e.g., Kravis 1984). 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 in 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 (CPD) method 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. (A bridge country is one that links together two countries through a comparison of each with the bridge country.) The CPD and EKS methods generally produce similar results (Krijnse-Locker 1982).

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 (πs), one for each basic heading. Each average price (π_i) is calculated from the category PPP for each of the n countries weighted by the quantity share 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,

$$\pi_i = \sum \frac{\mathrm{PPP}_{ij}}{\mathrm{PPP}_j} \cdot w_{ij} \tag{2}$$

where i is one of m categories or basic headings, and w is country j's share in world consumption of category i.

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

$$\mathrm{PPP}_j = \frac{\sum^{m} \mathrm{PPP}_{ij} Q_{ij}}{\sum^{m} \pi_i Q_{ij}} \tag{3}$$

Thus the PPPs for GDP (or other aggregates) need the πs for their calculation, and the πs need the PPPs 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 πs 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, the formula matches the underlying point of departure of the PPP comparisons—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 mea-

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, $I_{jk} = I_{jl} \div I_{kl}$ 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.

suring 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 eighty 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 4 study (eight countries) was only half of that of the United States; for the developing countries of Central and South America, the price level was less than two-thirds that of the United States. 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.

12.3 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. These uncertainties in the estimates of each country's national income pose the same problems for exchange-rate conversions as for PPP conversions. The seemingly unequivocal character of exchange-rate–based comparisons involves the conversion of these incomparable measures into equally incomparable estimates in the numeraire currency.

There are, however, other sources of uncertainty in the PPP conversions, and we try here to evaluate them.

12.3.1 Price Comparisons

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 Community (EC) meet these problems in a particularly thorough and careful way. The price comparisons involving the United States for the Phase 4 study were less than optimal because the decision of the United States 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 diffi-

cult 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. 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 and errors in matching are also difficult to measure. An experiment reported in the Phase 1 study (Kravis, Kenessey, Heston, and Summers 1975, pp. 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 was the most recent study offering much relevant material for this purpose).

12.3.2 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 Communities 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 United States 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 3 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 (with real GDPs per capita 15 percent or less of the U.S. level) were smaller only by a range of 9 to 13 percent.

Exchange-rate–converted per capita GDPs were, by contrast, 60 percent 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 3 study, the Geary-Khamis and EKS formulas produced nearly identical results, except for Japan, for which Geary-Khamis was 5 percent higher. The spread was larger for middle-income countries and greatest for low-income countries, averaging around 16 percent for the lowest fourth and reaching as high as 19 percent.[10]

12.3.3 Treatment of Problem Categories

For certain categories, there is little theoretical guidance for the choice between alternative methods.

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, produced the comparisons, and the differences in methods resulted from efforts to refine and improve the treatment of some difficult categories. In later phases of the ICP, the design and collection of the ICP was organized on a regional basis (more on regionalization presently), and differences arose even within a given phase.

Aside from regionalization, differences related mainly to the treatment of the net foreign balance and about a dozen "comparison-resistant" service categories for which domestic national accounts generally use input data to measure output.

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; Organisation for Economic Cooperation and Development, U.N. Statistical Office 1989.

10. These calculations are based mainly on the thirty-four Phase 3 countries. An increase in the number of countries studied might turn up some larger differences.

Comparison-Resistant Services

The proportion of GDP spent on services ranges from less than 20 percent in low-income countries to 50 percent 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 composed of the services produced by health care professionals, teachers, and government employees. They are "comparison-resistant" because it is difficult to find markets on which units of these service outputs 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 United Kingdom with thirteen or fourteen years of education. Exceptions to this equal productivity assumption were made in Phase 3 and for the portion of Phase 4 comparisons involving Austria and the three eastern European countries.

Another difference in the treatment of comparison-resistant services was that allowances were made in phases 1, 2, and 3, 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 4 and 5 of the ICP abstained from their use. The result of omitting them, however, is surely an understatement of output in high-income countries, since both health 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 inclusion in Phase 3 of the number of students, in addition to the input of teachers and capital. The change may have been introduced with the idea of adding a measure of the output of education, but, because students' time is an input into the production of learning (learning itself being the output), the procedure of Phase 3 represents an enlargement of the coverage of inputs rather than a measurement of output. 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.

11. Based on thirty-four countries (Kravis, Heston, and Summers 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.

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; quite a lot for low-income countries at the sector level; and modest amounts for low-income countries at the GDP level.

What is probably a maximum estimate of the sensitivity of GDP to these variations 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 3 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 percent, but would have little effect on the middle- and high-income countries.

The impact of different treatment of the comparison-resistant services is more realistically measured by examining the results of the Phase 3 changes in making comparisons for these services relative to those used in Phase 2. The main changes were the use of capital inputs in health and government services, an adjustment for productivity differences in health and education services, and the addition 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 United States (Kravis, Heston, and Summers 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 1, 3, and 4, using as nearly as possible the same methods in all three.[12] The methods were mainly those of Phase 4, the chief exception being that the fixity rule (see section 12.4) was not adhered to. The "standardized" results compare with the original benchmark results as shown in table 12.1.

The differences produced by the shift to alternative methods are quite small on average; ignoring signs, the mean difference varies from 4 percent to a little over 6 percent for the three benchmark surveys. The differences for individual countries in a few cases are as high as 17 percent, with an outlier at 25 percent (Malaysia, in 1970).

The Net Foreign Balance

The great preponderance of GDP is absorbed by domestic spending on consumption, capital formation, and 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

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

Table 12.1 **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
Number 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

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 EC and used in the more recent phases of the ICP, is simply 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 means of equation (2) above. A closer approximation to this method, involving the estimate of a π for the net foreign balance, was used in the first three phases of the ICP.[13]

The difference in the estimate of per capita GDP from these two ways of handling the category depends on the size and sign of the net foreign balance in each country and on the size of the π. In the Phase 3 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 was Zambia's, -19.5 percent of its GDP in its own currency. If the conversion to international dollars in the Phase 3 report had been at a price of \$I = \$1.00 instead of \$I = \$1.28, Zambia's per capita GDP relative to the United States 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

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 4 and 5, in which the net foreign balance is not multiplied by an international price, is equivalent to taking that price as equal to one.

foreign balance are likely to have modest effects on comparative real GDP per capita, even if the surplus or deficit is very large.

12.3.6 Robustness Summary

The main concern that has been expressed about the benchmark results is that they exaggerate the relative real per capita GDPs of the low-income countries. Uncertainties may arise from measurement error, improper matching of qualities, sampling variability, and the treatment of problem categories. We do not attempt to measure either the errors common to exchange-rate and PPP comparisons or the effect 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 uncertainty to get some rough approximations of the possible variation.

12.4 The Range of Uncertainty

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 allowance suggested above for sampling variability and for alternative methods. The worst case is that the true estimate will be approximately 10 to 15 percent 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-resistant 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 applied as uniformly as possible to the data of the 1970, 1975, and 1980 phases. In this approach, errors or uncertainty arising from methodological factors may raise some estimates *above* the benchmark estimate. The worst case observed in table 12.1 suggests that the methodological factors could place the true per capita GDP 17 percent higher than the benchmark estimate. If sampling errors are in the same direction, the up-side margin could be approximately 25 percent.

We conclude that margins of uncertainty in the 20 to 25 percent 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 percent, 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 about 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 multiple rates apply to different transactions, or, more commonly, black market rates exist along with fully legal rates, then 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 for 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 fluctuate 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 from "equilibrium" 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 adjacent years, adjusting the latter rates to the target year by relative rates of inflation.

The differences produced by these and other alternative exchange-rate methods are substantial. Ward (1989), for example, reports on the results 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 single source of different methodology produces differences in exchange-rate conversions that are larger than those encountered in the PPP-converted comparisons.

12.5 The Question of Regionalization

In accordance with long-held plans, beginning with Phase 4 (reference date 1980) the responsibility for worldwide comparisons was shifted completely to the U.N. Statistical Office (UNSO). At the same time, outside financial support dwindled, and the ability of 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 sixty countries were put together in seven sets of countries. In Europe, for example, the EC made up one block of twelve countries, and five 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 Organisation for Economic Cooperation and Development (OECD). The UNSO took direct responsibility for seven Asian countries for which no regional organization formed a comparison 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).[14] 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 twenty core countries for each basic heading, using the United States as the numeraire country.[15] The other (noncore) 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 sixty countries included in Phase 4.

This Geary-Khamis calculation would have provided the final result except for the insistence of the Statistical Office of the European Communities (Eurostat) on the "fixity" principle. Under this rule, the results of the intraregional comparisons produced by the different regional organizations are not to be altered when the regions are incorporated into the worldwide comparisons. That is, if Germany is 5 percent higher than France in real per capita income

14. 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; some "regional" groups cut across continental lines (OECD) and others subdivide continents (EC) (Kravis, Heston, and Summers 1982).

15. The CPD method was used.

in the EC calculation, the difference must be maintained when the EC countries are put into the context of OECD comparisons or worldwide comparisons. The strong support of the fixity principle by Eurostat was based on 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 sixty-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 1–3) and twenty-three additional aggregations. This distribution was based on each country's distribution as produced by the intraregional comparison. This method of integrating the results of different regional comparisons (the "GDP consistency" method) has the virtue of comparability for the quantities of GDP at world prices (international dollars). 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 worldwide comparison. For these subdivisions of GDP, the results are not comparable for countries in different regions because they are based on regional rather than world prices. Very large distortions, some in excess of 100 percent, 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 4 and 5 comparisons at worldwide average prices. Also the restrictions imposing fixity are to be lifted about three years after publication of the first regional results. Since Eurostat has put out its results well before 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 Germany-Japan comparison, for example, might depend on average prices

based on EC, European, Asian, and world prices, while the Germany-France comparison might be based on European prices alone, and the Germany-United States, on European, North American, and world prices.

This is not to say that regional comparisons should not be made. Their 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.[16] For some purposes 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 EC, a community-wide set of average prices would be more appropriate than either an average of French and German prices or average world prices, the latter including the price structures of such diverse countries as the United States and India. For general-purpose comparisons, however, the use of world average prices seems most appropriate. 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 EC officials and politicians that the EC comparisons are best for EC purposes even though another set exists.

12.6 Extensions to Nonbenchmark 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 less-developed countries (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 fifty 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. A number of ways have been suggested to prepare estimates of comparative real GDP per capita for missing countries by means that entail a smaller ex-

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 sixty countries of Phase 4 (1980) published by 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 twelve other countries and over 10 percent for eight others (the largest difference was 21 percent).

penditure of resources (see Heston 1973). These various approaches are described and illustrated here.

The method closest to the benchmark studies, the "reduced information" approach, carries 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 post-adjustment price comparisons that will yield approximately the same estimate as the benchmark comparisons.

Several other methods are 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 both 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 nonbenchmark country is obtained by plugging into the estimating equation the values of the independent variables for that country.

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; only the latter predictions are reported here. (Virtually all nonbenchmark countries were in the developing category.) The equations were used to estimate domestic absorption, and GDP was then obtained by adding the net foreign balance, which was converted to dollars via the exchange rate (see Summers and Heston 1984).

12.6.1 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 thirteen developing countries by Sultan Ahmad of the World Bank (Ahmad, 1988). Ahmad began by experimenting with ICP Phase 3 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 subsequent estimates for the countries in his experiment, summarized in table 12.2, range from 68 percent of the benchmark estimate to 103 percent. The mean absolute deviation is around 10 percent. There is a tendency for the deviation from the benchmark to be smaller for the higher-income countries. In every case the reduced information estimate is closer to the ICP benchmark than is the

Table 12.2 **Estimates of Per Capita GDP by Reduced Information Method Compared with Benchmark Estimates, 1980 (U.S. = 100)**

				Ratio to Benchmark	
	Exchange Rate Converted (1)	Reduced Information (2)	Benchmark (3)	Exchange Rate Converted (1) ÷ (3) (4)	Reduced Information (2) ÷ (3) (5)
Costa Rica	18.5	21.3	27.2	.68	.78
Dominican Republic	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
United States	100.0	100.0	100.0	1.00	1.00

Source: Ahmad (1988).

exchange-rate converted GDP. Ahmad finds his reduced information estimates also come closer to matching the benchmark estimates than short-cut estimates produced by Summers and Heston (1988) 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 poor countries for which benchmark studies cannot be readily carried out or for which the costs would be too high.

12.6.2 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).

In the physical indicator approach, the relationship between real GDP per capita and each of a score or more of indicators is examined (see, e.g., 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 multicollinearity 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 cut-off 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]

12.6.3 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 may include openness, price isolation, money growth, and the trade balance (Kravis, Heston, and Summers 1978b, Kravis and Lipsey 1983, Clague 1986). Education and shares of minerals in GDP are sometimes included in this approach, as they are in the physical indicator method.

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. The coefficient in this relationship is not 1; 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 ratio of r to n falls as n increases. Although the line of causation presumably runs from r to n, r is treated as a function of n for purposes of extrapolation.

12.6.4 Short-cut Methods: Price Indicators

Many multinational business enterprises and international organizations have employees stationed in different countries with different price levels and encounter the need to maintain a system of "post-adjustment" allowances to equalize the real incomes of personnel of equal status in the headquarters locations with those stationed elsewhere. This work often 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 Labor 1981) and the United Nations for about 125 cities (U.N. 1980). Private organizations also

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. Thus, in this procedure the estimating equation used for real GDP per capita varies from one country to another.

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. Estimates for nonbenchmark 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 is the same everywhere between the national price level and the price level encountered by foreign personnel dwelling in the capital or another leading city for professional or business reasons. Experiments with alternative sources of post-adjustment price comparisons showed that they produced very similar results. For brevity and simplicity, we report here only on the results using the U.N. data.

12.6.5 A Comparison of the Results of the Short-cut Experiments

In table 12.3, some results of shortcut methods are presented; the shortcut estimates are expressed as ratios of the benchmark estimates. For comparative purposes, the exchange-rate conversions are also presented. Underlying equations appear in the appendix, table 12A.1.

Table 12.3 **Predictions of LDCs 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 Deviation
	Maximum	Minimum	
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

Source: See appendix table 12A.1.

Note: The phase 4 countries were arrayed according to increasing real GDP per capita as measured in the benchmark study and were 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 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.

The basic procedure was to array the developing countries in the 1980 (Phase 4) comparisons 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 this 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 that it is not dependent on exchange-rates and related prices. Its predictions are on the average within 12 percent of the benchmark estimates, although the range is from predictions 46 percent above the benchmark to predictions as much as 32 percent below.

All the short-cut estimates clearly outperform the exchange-rate conversions; the latter are characterized by mean deviations of 30 percent and understatements of relative GDPs by more than 70 percent.

12.7 Extensions to Other Years

The availability of benchmark studies for quinquennial years leaves open the question of estimates for the in-between years. 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 is possible of the benchmark-year estimate for real GDP per capita relative to that of the numeraire country. 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 United States 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 benchmark-year international prices.

A question that remains 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 do 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, and so forth. Extrapolated values would be filled in for those components not covered in the pricing of that year. This has the merit of integrating ICP data with the price and quantity 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 valuable 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 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 over 130 countries annually from 1950 to 1988 has been offered by Robert Summers and Alan Heston (1985, 1991). 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.[18] They show breakdowns for consumption, investment, and government both in current prices and 1980 prices and also in the price level 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.[19]

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

$$\frac{y_{t+5}}{y_t \cdot g} = 1$$

18. For some users of these comparisons, presentation of the original benchmark data in the 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*).

19. For an analysis based on a different concept of world prices, see Bhagwati and Hansen (1973).

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); y_t+_5 is the same for a benchmark five years later (1980); and g is the 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 percent, and the largest increase about 9 percent. 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 that would often require data not in the public domain—is that g is based on a mixture of domestic and international prices. The reason is that for practical reasons (the unavailability of price indexes) the extrapolation of y_t in 1980 has to be done, 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. 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. Then an average of the two is taken. This process introduces ambiguities about the date of the price structure that is being applied.

Even if this problem were to be corrected, the massaging of the benchmark comparisons and of the growth rates 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 compensates 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.

Our view is that the best general-purpose estimates of growth rates are those

derived directly from the national accounts—from 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 Summers and Heston have produced the most comprehensive set of PPP-based estimates that exists; their "consistentized" data cover almost all countries and thirty-five years. As noted, their data are aimed at uses requiring consistent estimates of levels and of changes in output.

12.7.1 Growth Rates

The usual way of calculating growth rates is to take the changes in the real GDP of each country as measured using its own market basket and its own base-year prices. Thus, the growth rate of a given country measures the change in a basket of goods that is different from the basket 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 1, compared to the change in the quantity of the different base (current) bundle produced in country 2?" 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 drawback 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 1, compared to the change in the total quantity of goods absorbed in country 2, recognizing that the list of goods may be different in each situation, but valuing the goods at 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;

for the world as a whole, the international prices show the rates of transformation of different goods.

12.8 Future ICP Work

Phase 6 (reference date 1990) is encountering difficulties because a number of the binary comparisons which were to be used to link regions appear to be languishing. There is a great danger that Phase 6 will be a series of regional comparisons without enough links to produce systematic worldwide comparisons. A way should be found that does not require large resources to revive the prospects for worldwide comparisons.

One possibility is to modify the comparison strategy from a core-country approach to a core-commodity approach. The U.N. Statistical Office would develop a core-commodity list consisting of items found in most if not all regions (e.g., eggs, sandals). Countries would be asked to provide prices for as many as possible of the core items. In addition, each country would price items that were specific to its own region. Between the two lists of specifications, there would be an adequate number to make possible price comparisons for each basic heading. The price comparisons between the regions would be based on the common core. It would not be necessary for every country to price every core commodity. The price comparisons between countries within a region would be based on a combination of core commodities and commodities found mainly or only in the region.[20]

As already suggested, a consequence of the reliance on regions to organize much of the data in phases 4 and 5 has been the emergence of some differences in methods, although the broad strategy of the ICP and most of its detailed methods were adhered to.

The differences are greater than they would have been had UNSO the resources to coordinate the work of the different regions. The differences will grow larger unless UNSO takes a more extensive role. U.N. 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 could be matched in other regions and to provide UNSO with data to 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 UNSO with the standard data-set, they should be free to use differing methods in their own regional comparisons.

20. CPDs could be calculated for each region and used to fill in missing prices in the regional set. Then a second-stage CPD could 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 could be the inputs for a Geary-Khamis calculation 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).

Another important role that a U.N. presence at planning meetings might fill 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 nonbenchmark countries.

12.9 Conclusion

A system of international comparisons of income and of the purchasing power of currencies is now in place, covering most countries and currencies for the period 1950–85. 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 eighty 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 worldwide 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 composing each country's national absorption of final goods (and net claims against foreigners). Given this approach, the benchmark results are not very sensitive to plausible alternative treatments of certain methodological issues, to the resolution of which theory gives little guidance. ("Plausible" alternatives include those seriously considered or adopted by UNSO, Eurostat, ECE, or OECD.)

We conclude that margins of approximately 20 to 25 percent 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 percent.

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.

Even if these estimates of the range of uncertainty 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; for the group of developing countries the average ratio is over two times. The exchange-rate converted figures are farther from the lowest PPP-converted estimates of real GDP per capita that emerge from the considerations concerning uncertainties.

The uncertainties of PPP conversions are inherent in international comparisons. They can be reduced by the investment of further resources, especially by extending benchmark estimates to more countries (particularly low-income ones) or years. Time may bring further consensus on methodological questions and narrow the uncertainty range. But it is unrealistic to expect that real product estimates will ever have the illusory certainty of exchange-rate converted estimates. Exchange rates, however, can be justified as converters only 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. Users of real GDP comparisons can either delude themselves with unequivocal but wrong—often far wrong—and biased estimates of real GDP based on exchange rates, or accept the fact that the closest we can come to comparative GDPs involves uncertainties about the exact figures.

The future of ICP estimates appears assured in Europe, particularly in the EC. In other regions, prospects vary, but as of mid-1990 the outlook for systematic worldwide comparisons for Phase 6 (1990) does not look bright. It will take a renewed impetus, which in the circumstances can only be provided by 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 the momentum that has been gained with great effort toward the final step in establishing comparability—the translation of own-currency GDPs into comparable measures of real income.

Appendix

Table 12A.1 Equations Used to Generate Predictions Summarized in Table 12.3

I. *Physical Indicators*

A. $DA_{pc} = .763 + .193$ (calorie intake p.c.) $+ 1.698$ (life expectancy)
(0.36) (1.56) (2.82)

$+ .225$ (energy consumption p.c.) $- .429$ (% of labor force in agriculture)
(2.95) (3.90)

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

B. $DA_{pc} = -3.899 + .543$ (calorie intake p.c.) $+ .201$ (secondary school)
(1.42) (1.74) (2.43)

$+ 1.501$ (life expectancy) $+ .194$ (energy consumption p.c.)
(3.21) (3.67)

$- .237$ (% of labor force in agriculture)
(4.17)

$\bar{R}^2 = .97$

II. *Price Indicators*

A. $DA_{pc} = 2.18 + .744\ (DA_{UN}) - .221\ (D_A)$
(8.46) (23.16) (2.44)

$\bar{R}^2 = .97$

B. $DA_{pc} = 2.448 + .742\ (DA_{UN}) - .397\ (D_A)$
(11.25) (27.02) (5.44)

$\bar{R}^2 = .98$

III. *Monetary Indicators*

A. $DA_{pc} = 1.305 + .691\ (n) - .036$ (openness) $- .442\ (D_A)$
(9.79) (21.17) (0.54) (4.64)

$\bar{R}^2 = .97$

B. $DA_{pc} = 1.641 + .620\ (n) + .129$ (openness) $- .649\ (D_A)$
(12.75) (19.61) (1.61) (6.93)

$\bar{R}^2 = .97$

Notes: The predictions for developing-country real GDP per capita that are generated by these equations are summarized in table 12.3. The regressions are based on all the Phase 4 countries and are in log form; *t*-statistics are 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 U.N. post-adjustment data; D_A = dummy variable with value of 1 for African countries; O for others; n = exchange-rate–converted GDP per capita; openness = 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. *t*-ratios are in parentheses.

References

Ahmad, Sultan. 1988. International Real Income Comparisons with Reduced Information. In *World Comparisons of Incomes, Prices, and Product*, ed. J. Salazar-Carrillo and D. S. Prasada Rao. North Holland: Elsevier Science Publishers.

Beckerman, Wilfred. 1966. *International Comparisons of Real Incomes*. Paris: Development Center, Organisation for Economic Cooperation and Development.

Bhagwati, Jagdish N. 1984. Why Are Services Cheaper in Poor Countries? *Economic Journal* 94(374): 279–86.

Bhagwati, Jagdish N., and Bent Hansen. 1973. Should Growth Rates be Evaluated at International Prices? In *Development and Planning*, ed. J. Bhagwati and Richard S. Eckaus.

Clague, Christopher. 1986. Short Cut Estimates of Real Income. *Review of Income and Wealth* (September): 313–31.

Drechsler, Laszlo. 1988. The Regionalization of and Other Recent Developments in the U.N. International Comparison Project. In *World Comparisons of Incomes, Prices, and Product*, ed. J. Salazar-Carillo and D. S. Prasada Rao. North Holland: Elsevier Science Publishers.

Gilbert, Milton, and Irving B. Kravis. 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: Organisation for European Economic Cooperation.

Heston, Alan W. 1973. A Comparison of Some Short-Cut Methods of Estimating Real Product Per Capita. *Review of Income and Wealth* 19(1): 79–104.

Hill, T. Peter. 1977. On Goods and Services. *Review of Income and Wealth* 23(4): 315–38.

———. 1981. Multilateral Measurements of Purchasing Power and Real GDP. Report prepared for the Statistical Office of the European Communities and the Economic Commission for Europe. Luxembourg.

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

Isenman, Paul. 1980. Inter-Country Comparisons of "Real" (PPP) Incomes: Revised Estimates and Unresolved Questions. *World Development* 8(1).

Kravis, Irving B. 1984. Comparative Studies of National Incomes and Prices. *Journal of Economic Literature* 22 (March).

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

———. 1978b. Real GDP Per Capita for More Than One Hundred Countries. *Economic Journal* 88(350): 215–42.

———. 1982. *World Product and Income: International Comparisons of Real Gross Product.* Baltimore: Johns Hopkins University Press.

Kravis, Irving B., Zoltan Kenessey, Alan W. Heston and Robert Summers. 1975. *A System of International Comparisons of Gross Product and Purchasing Power*, Baltimore: Johns Hopkins University Press.

Kravis, Irving B., and Robert E. Lipsey. 1983. Toward an Explanation of National Price Levels. Princeton Studies in International Finance, no. 52 (November).

Krijnse-Locker, Hugo. 1984. On the Estimation of Purchasing Power Parities on the Basic Heading Level. *Review of Income and Wealth* 30(2): 135–52.

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. Typescript.
Paige, D., and G. Bombach. 1959. *A Comparison of National Output and Productivity of the United Kingdom and the United States*. Paris: Organisation for European Economic Cooperation.
Ruggles, Richard. 1977. *The Wholesale Price Index—Review and Evaluation*. Report of Council on Wage and Price Stability (June). Washington, D.C.
Statistical Office of the European Communities. 1977. *Comparisons in Real Values of the Aggregates of ESA, 1975*. Luxembourg: European Economic Communities.
Statistical Office of the European Communities, Organisation for Economic Cooperation and Development, United Nations Statistical Office, *Expert Group on ICP Methodology,* June 14–16, 1989) Paris; and *Report of the Expert Group on ICP Methodology,* (June 6–10, 1988) Luxembourg.
Summers, Robert. 1973. International Comparisons Based upon Incomplete Data. *Review of Income and Wealth* 19(1): 1–16.
Summers, Robert, and Alan W. Heston. 1984. Improved International Comparisons of Real Product and Its Composition, 1950–80. *Review of Income and Wealth* series 30, no. 2 (June): 207–62.
———. 1988. A New Set of International Comparisons of Real Product and Price Levels: Estimates for 130 Countries, 1950–1985. *Review of Income and Wealth* (series 34, (March): 1–25.
———. 1991. The Penn World Tables (Mark V): An Expanded Set of International Comparisons, 1950–88. *Quarterly Journal of Economics* (May): 327–68.
Summers, Robert, Irving B. Kravis, and Alan W. Heston. 1980. International Comparisons of Real Product and Its Composition, 1950–77. *Review of Income and Wealth* 26(1): 19–66.
United Nations. *Monthly Bulletin of Statistics*, 1975 (August), 1980 (September).
U.N. Economic Commission for Europe. 1980. *Economic Bulletin for Europe* 31(2)
U.N. Statistical Office. 1968. *A System of National Accounts*. Studies in Methods, series F, no. 2, rev. 3.
U.S. Department of Labor. 1981. United States Department of State Indexes of Living and Costs Abroad and Quarters Allowances. Report 646 (April).
Ward, Michael. 1985. *Purchasing Power Parities and Real Expenditures in the OECD*. Paris: OECD.
———. 1989. The Reliability of Estimates of Per Capita Income. Paper prepared for conference of the International Association for Research in Income and Wealth. August 20–26.

Comment

Alan V. Deardorff

What a pleasure it has been to read this paper. I have had a passing awareness of the International Comparison Program, of course, for many years. But I have never had occasion to make direct use of the ICP data, as many others here have, and I therefore never came to learn in detail how it had all been put together. This paper does a fine job of introducing the ICP, and I would have

Alan V. Deardorff is professor of economics and public policy at the University of Michigan.

learned a great deal from it for that reason alone. The full picture of the project can be seen in the various reports, mainly by Kravis, Heston, and Summers, that are referred to in the paper. I will comment mostly on the current work dealing with robustness.

One of the unintended messages of the paper, of course, is that the ICP has entailed a huge amount of painstaking work, over many years and by many individuals and agencies. I can't imagine anyone voluntarily getting into such a project knowing in advance how much work it would be and the difficulties and complications that would arise. Were it not for the fact that the authors have been involved in other projects as well that are of daunting size and complexity, I would have assumed that they did it unintentionally. However, since this is their second paper in this conference reporting on such work, I have to conclude that they are economists of exceptional courage, or perhaps masochism.

But I must be careful, or I will give the impression that this is a paper that requires an equal dose of courage or masochism to read. That is not the case at all. The massively detailed body of data that the ICP has produced over the years is very much behind the scene here. The paper itself, based on the various Kravis, Heston, and Summers reports, is a wonderfully careful and thoughtful discussion of the methodological issues that have come up along the way. It is clear that over the years the authors have confronted numerous problems with the data and its interpretation, as well as conceptual problems as to how to make valid international comparisons. They share with us here what these problems have been and how they have been resolved.

Finally, and this is the unique contribution of the Kravis and Lipsey paper presented today, they have been able in a surprising number of cases to quantify the limits of error that may have been introduced by the choices they have made or been forced to make. In a conference that deals with the limitations of available data, the paper is a prototype of how these limitations should be acknowledged and evaluated in assessing the results that are based upon those data. We would all do well to follow their example in this regard, though few of us will have two decades of research on a single topic to draw upon.

This conference has dealt primarily with issues of data, and seems therefore to be concerned primarily with empirical issues rather than theoretical ones. Yet the thoughtful approach of this paper shows, I think, that the very inadequacies of the data provide a need for more careful theoretical research as well. The ICP staff has shown wisdom and ingenuity in dealing with these inadequacies, without apparently much guidance from the theoretical literature. Those who follow will do well simply to follow their example, but it would be useful if these issues could be dealt with in more general terms, so that empirical research using inadequate data could in the future be guided by general principles rather than have to develop solutions to problems as they arise, as Kravis and Lipsey have done here. I would therefore like to enter a

plea for both economic and econometric theorists to devote more of their theoretical attention to the problems of inadequate data and the appropriate methods for dealing with them.

Edward Leamer, in his comments on Keith Maskus's paper, gives us a list of many of the problems that arise with data, especially in the international context. It is fascinating to see how many of these problems have come up in the International Comparison Program and been dealt with there.

Consider first the problem of missing data. The ICP is plagued by missing data at the most basic level, since many products are not produced in common across all countries. Yet prices of these goods must nonetheless be compared across countries. In their other paper on prices, Kravis and Lipsey mentioned the Country-Product Dummy (CPD) method of constructing missing prices. This method was actually developed by Robert Summers specifically for use in the ICP.

The ICP produces a full comparison of prices and real outputs only for selected years and countries. This leaves as missing data the values for the intervening years and excluded countries. The ICP has used a variety of methods that are detailed in the paper for filling in these missing values. Like the CPD method for constructing missing prices, these methods also are based on regressions of known values on various explanatory variables that are somewhat ad hoc but nonetheless seem to do a decent job of predicting the missing values.

What I would like to see is a more general treatment of this problem of missing data. I understand that the econometrics literature has dealt with how best to perform estimation when there are missing observations, but this literature is directed at perfecting the estimation, not at replacing the missing observations themselves. Yet it strikes me that there must be some general guidance that could be provided on this issue, and that very likely such an analysis would lead one to perform very much the sorts of calculations that Kravis and Lipsey have in the ICP.

My point is not that this kind of analysis is necessarily needed for the ICP; it has done a more-than-adequate job of dealing with these issues. Rather, there are many areas of empirical international economics (and no doubt of empirical economics more generally) where missing data are a recurring problem, and a systematic methodology for filling them in would be very useful. I know, for example, that in the work Robert Stern and I do with computational general equilibrium models, missing data are a constant source of difficulty.

Edward Leamer also listed the problems caused for international data by exchange-rate gyrations and the more general issue of noncomparability across countries of international data. These, of course, are precisely the problems that the ICP has dealt with all along. Most of us are content to use the most readily available data, perhaps acknowledging its deficiencies, perhaps not. Kravis and Lipsey and their coauthors have confronted these problems head-on and have sought to correct them—with remarkable success!

Here, incidentally, the problem is not one of missing data, or even necessarily of bad data. Rather, it is that the data one seeks in some sense *cannot* exist. There is no uniquely correct meaning that attaches to the concept of real GDP, given that countries differ as they do in the assortments of goods and services that they produce and in the prices at which they produce them. The problems are familiar index-number problems, and so today are not a very hot topic in economic theory. Yet the contortions that have been endured in dealing with them in the ICP indicate to me that the conceptual problems have not all been solved. The ICP relies on something called the Geary-Khamis method of aggregation, for example, and Kravis and Lipsey make a compelling case in its favor in terms of various plausible and apparently desirable properties that it possesses. Yet I would have thought that this kind of problem could be addressed from a theoretical standpoint that, though perhaps not conclusive, could nonetheless shed light on what is and is not being measured.

I'll conclude with a question for the authors that may not be fair, since it deals with something that was not in the paper. One of the most important findings of the ICP is that developing-country GDPs are routinely underestimated by converting at nominal exchange rates. This means that the nominal values of LDC currencies are typically quite a bit below their PPP levels. This finding intrigued me because I thought I had understood LDC currencies to be typically overvalued, not undervalued. If the findings were true, it would merely underscore the well-known inadequacy of PPP as a guide to equilibrium exchange rates. Yet the authors told me that they doubted that it *was* true. Indeed, they said that, though not reported in this paper, the ICP has found that even traded-goods prices in LDCs are lower at prevailing exchange rates than comparable prices in developed countries, and that the result for real GDPs is therefore not just a reflection of exceptionally cheap nontraded goods, as I had supposed.

Now this surprises me a great deal. For whether or not LDC exchange rates are overvalued, it is certainly the case that most LDCs have high barriers to trade, both tariffs and nontariff barriers, and it is my impression that they also have frequent subsidies to exports as well. Both of these will lead to the prices of imported and exported goods being higher within the LDCs than on world markets. If indeed the prices of domestically produced tradable goods for domestic absorption are lower in these countries than abroad, then it must surely suggest another source of noncomparability.

In computable general equilibrium (CGE) modeling we routinely find that traded goods must be modeled as imperfect substitutes for apparently comparable traded goods from abroad. Otherwise it is impossible to replicate the data on the amounts and price-responsiveness of trade. I had supposed that this imperfect substitutability, which is usually modeled via the Armington assumption of goods that are differentiated by country of origin, was forced upon us by our levels of aggregation. But if the ICP finds such price discrepancies even at the very disaggregated level where they do their price compar-

isons, then I wonder if in fact the goods produced in LDCs are not after all rather imperfect substitutes for their developed-country counterparts.

If so, and if this by any chance means that the LDC products are of lower quality overall than other goods, then the conclusion that LDC real GDPs are understated may be suspect.