

# Alternative measures of output in global economic-environmental models: Purchasing power parity or market exchange rates?

William Nordhaus <sup>\*,1</sup>

*Yale University, USA*

Accepted 7 February 2006  
Available online 6 March 2006

---

## Abstract

The study examines the question of the use of purchasing power parity versus market exchange rates in constructing global economic models. It compares three approaches: MER accounts, world-price PPP accounts, and superlative PPP accounts. It concludes that the best approach is to use superlative PPP accounts. This approach uses cross-sectional PPP measures for relative incomes and outputs and relies on national accounts price and quantity indexes for time-series extrapolations. Under ideal circumstances, this approach will provide accurate and consistent cross-sections and time series. This approach will require relatively little change in model structure from ones using MER accounts. The main concern would be to ensure that behavioral and reduced-form relationships in the models have been correctly estimated. © 2006 Elsevier B.V. All rights reserved.

*Keywords:* Global environmental models; Climate change; Purchasing power parity; Market exchange rates

---

Most global energy-economic-environment models constructed over the last three decades have relied on market exchange rates (MER) to put countries in a common metric for estimation and calibration. This approach has been the subject of considerable discussion in recent years, and the alternative of purchasing power parity (PPP) exchange rates has been proposed. The present analysis reviews the issues and describes a reconciliation of the debate.

---

\* 28 Hillhouse Avenue, New Haven, CT 06511, USA.

*E-mail address:* [william.nordhaus@yale.edu](mailto:william.nordhaus@yale.edu).

<sup>1</sup> This report was prepared for the IPCC Expert Meeting on Emission Scenarios, US-EPA, Washington, DC, January 12–14, 2005.

### Alternative Accounting Systems

	Market exchange rates (MER)	Purchasing-power parity exchange rates (PPP)
Actual prices for all goods and services	A. MER accounts: incorrect estimates of income and output	C. Superlative PPP accounts: correct income and price behavior
World prices for all goods and services	No sense.	B. World-price PPP accounts: correct incomes but inconsistent with market behavior

Fig. 1. Alternative approaches to constructing international economic models.

The plan of this report is the following. After an overview, the first section reviews the merits of different systems for measuring real income and output. The second section reviews current practice in constructing comparative economic accounts, with particular attention to construction of PPP accounts. The third section discusses the quality of the data in different systems. The fourth analyzes the relative merits of different price measures in model construction, after which issues of convergence are addressed. The sixth section then surveys the principal arguments regarding different accounting approaches to constructing global models. The last section summarizes the paper. Three appendices provide further detail on estimates of growth rates in different databases, concepts of superlative index numbers, and issues of aggregation of regions into larger aggregates.

Before beginning, we need to settle on some linguistic conventions. First, “PPP accounts” refer to accounts that estimate aggregate incomes, outputs, and expenditures correcting for differences in national price levels. The major obstacle to such accounts is accurate measures of prices of goods and services in different countries. Second, “world prices” refers to the synthetic average prices that are estimated in some PPP systems of accounts to estimate different national incomes or costs of living. “Actual market prices” are the local or national prices faced by market participants.

There are four different approaches to constructing comparative national accounts, as shown in Fig. 1. This matrix shows a two by two combination of the exchange rate convention and the disaggregated price convention. There are two possible approaches to converting incomes of different countries: purchasing power parity exchange rates (PPP) or market exchange rates (MER). Similarly, there are two different sets of relative prices: actual market prices or uniform world prices for individual goods and services. From this table, we see three serious candidates for accounting systems:<sup>2</sup>

MER accounts. The usual approach, shown in the upper left, uses actual market prices and market exchange rates.

<sup>2</sup> In an earlier draft, “superlative PPP accounts” were called hybrid accounts and “world-price PPP accounts” were called full PPP accounts. The terminology was changed to make the names more reflective of the actual designs.

World-price PPP accounts. Alternatively, we might use both PPP exchange rates and uniform world prices, as shown in the lower right entry.

Superlative PPP accounts. A third approach, shown in the upper right, uses PPP exchange rates and combines these with actual market prices for each country. These are called “superlative PPP accounts” because they rely upon superlative indexes over both space and time to compare outputs of different countries.

This study concludes that the best approach is to use superlative PPP accounts. It is “superlative” in the sense that it relies on consistent superlative indexes of output differences across space and time. To implement this approach, modelers should use cross-sectional PPP measures for relative incomes and outputs and national accounts price and quantity indexes for time-series extrapolations. Under ideal circumstances, this approach will provide accurate and consistent cross-sections and time series.

By contrast, estimates of output or income at MER are simply wrong — they are constructed on an economically incorrect basis. The world-price PPP measure is also undesirable because it uses prices that are not ones faced by firms and consumers and therefore does not accurately represent the detail of individual economies. I will explain each of these points in the sections that follow.

## **1. Which approach provides a more accurate measures of output and better measures of economic welfare?**

There has been a vigorous debate on the relative merits of PPP versus MER for many years, and recently this has been taken up the context of economic and emissions projections for climate change.<sup>3</sup> Much of the discussion on climate-change modeling revolves around specific projections (such as the “SRES” projections used by the IPCC analysis). Many of the points in the modeling debate are incidental to the fundamental issues of the appropriate accounting frameworks, so they will be largely outside the scope of the current review.

Which system provides more accurate measures of output, income, and economic welfare? In other words, when our analyses need to take into account the relative incomes and outputs of different regions, which system of accounts is preferred?

The basic answer is clear: In principle, PPP measures are superior to MER measures for representing relative incomes and outputs. (Strictly speaking, the “principle” in this assertion needs careful discussion, a point I return to in the next section.) We can show this point in Fig. 2

---

<sup>3</sup> There is an extensive discussion of the relative merit of different approaches to measuring incomes going back to early studies of Colin Clark, Milton Gilbert, and Irving Kravis. A comprehensive discussion is contained in Irving Kravis, Alan Heston, and Robert Summers, *World Product and Income: International Comparisons of Real Gross Product*, Johns Hopkins University Press for the World Bank, Baltimore, Md., 1982, available online at <http://pwt.econ.upenn.edu/papers/paperev.html>. In the context, of climate change modeling, MER was prominently criticized in Ian Castles and David Henderson, “Economics, Emissions Scenarios, and the Work of the IPCC,” *Energy and Environment*, vol. 14, no. 4, 2003, pp. 415–435. A summary response by a team of modelers is contained in Arnulf Grübler, Nebojša Nakićenović, Joe Alcamo, Ged Davis, Joergen Fenhann, Bill Hare, Shunsuke Mori, Bill Pepper, Hugh Pitcher, Keywan Riahi, Hans-Holger Rogner, Emilio Lebre La Rovere, Alexei Sankovski, Michael Schlesinger, R.P. Shukla, Rob Swart, Nadejda Victor, and Tae Yong Jung, “Emissions Scenarios: A Final Response,” *Energy and Environment*, vol. 15, No. 1, 2004, pp. 11–24. The extensive literature on this subject is largely incidental to the analysis here and is therefore not cited.

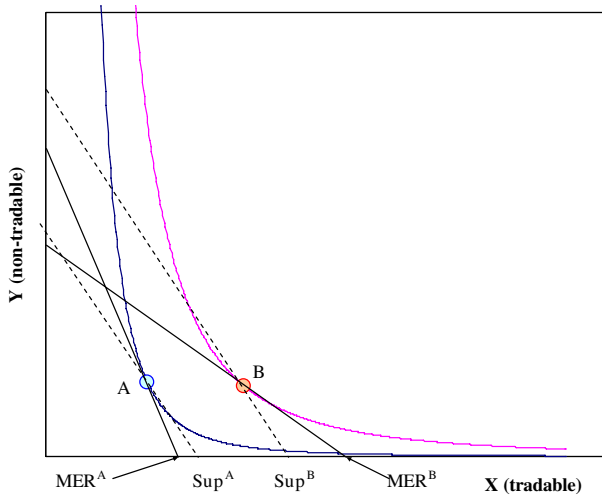


Fig. 2. Illustration of the bias in using MER. With two countries, we assume that the tradable goods are relatively more expensive in the poorer country. This shows that the output differential using market prices will be higher than when using a common price index or superlative output index.

using two countries and two goods.<sup>4</sup> The two countries (A and B) produce two goods (tradable goods on the horizontal axis and non-tradable goods on the vertical axis). Tradable goods are relatively more expensive in the poor country.

We can simplify our analysis by assuming that exchange rates are determined so that the prices of tradable goods are equal. We can then calculate total output in terms of tradable goods by taking the intersection of the price line with the horizontal axis. It is important to note that this provides the same estimate as output at market exchange rates. This shows that the output difference when measured at market exchange rates (which is the ratio  $MER^A/MER^B$ ) is relatively high.

Using MER will provide a misleading estimate of the differences in utility shown by the two homothetic indifference curves. Real incomes or relative utilities can be better approximated using superlative output indexes, shown as the dashes lines, with intersections on the tradable-goods axis at  $Sup^B$  and  $Sup^A$ . The output ratio ( $Sup^A/Sup^B$ ) is smaller than  $MER^A/MER^B$  because of the higher relative prices of tradable goods in the high-income country.

It should be emphasized that the use of MER is not a fine technical detail. Incomes estimated at MER are fundamentally wrong because they use the prices of a non-representative bundle of goods to compare the different countries. To use exchange rates is analytically no better than using the ratio of the prices of oranges or gasoline or haircuts to convert domestic incomes. A numerical example will illustrate the problem. Suppose that we decided to compare Iraq and Germany using the price of gasoline as the converter. Gasoline currently sells for \$0.05 per gallon in Iraq and \$5.00 per gallon in Germany. On this basis, per capita GDP in Iraq is \$78,000 as compared to \$25,000 in Germany. No one would take this example seriously as a description of the two economies, yet the use of market exchange rates has a similar basis.

The reason that using PPP-estimated income and output levels is empirically important lies in the finding that the use of MER underestimates current relative output levels of low-income countries by

<sup>4</sup> The figure is inspired by one in Steve Dowrick and John Quiggen, "True Measures of GDP and Convergence," *The American Economic Review*, vol. 87, no. 1, March 1997, pp. 41–64.

a factor of around three relative to high-income countries. This downward bias in MER is known as the Harrod–Balassa–Samuelson effect. This effect arises because tradable goods are currently relatively more expensive in low-income countries than in high-income countries, as shown in Fig. 1.

It might be tempting to take the current cross-sectional relationship between the MER/PPP ratio and per capita income and construct historical PPP output levels on that basis. This would be inadvisable for two reasons. First, the correct historical series should be built up from the (correctly constructed) national time series, as will be discussed in the next section. Second, it is unlikely that the current relationship is a universal constant. An important recent finding is that the strong positive relationship between per capita income and the MER/PPP income ratio did not hold uniformly over time. Indeed, there is evidence that the MER/PPP ratio–income relationship was slightly negative around World War I as compared to firmly positive today.<sup>5</sup>

## 2. Current practices for alternative measurement systems

I next turn to a discussion of the different accounting systems. Standard practice in global economic model building for climate change until recently has relied primarily upon national economic accounts converted by market exchange rates (MER).<sup>6</sup> While the construction of such accounts has evolved considerably, they are relatively easy to employ.

It has been argued that models would be more accurate if they were constructed using PPP exchange rates; it is unclear whether such recommendations also include using panels of PPP studies and world prices rather than national price data. This section reviews current practice in national economic accounting and purchasing power parity accounting.

### 2.1. Standard national accounts

National economic accounts with actual market prices have evolved greatly and are currently embodied in the System of National Accounts. The basic principle is to estimate values using observed or imputed market prices. Construction of price and quantity indexes generally assumes that they reflect the preferences of a “representative consumer.” Moreover, the preferences are generally assumed homothetic (constant-returns-to-scale), indicating that the relative preferences are independent of income and output. While the assumption of homotheticity is often rejected in the data, there are at present no realistic substitutes available for most countries (or possible for any country over a long time horizon).<sup>7</sup>

Increasingly, price and quantity indexes in national economic accounts are based on “superlative” indexes.<sup>8</sup> While the rigorous definition of this term is complicated, an intuitive description is that a superlative index provides a second-order approximation to a smooth homothetic function. There are many possible superlative indexes. One example is the Fisher ideal index, which is used in the

---

<sup>5</sup> The evidence is reviewed in Paul R. Bergin, Reuven Glick, and Alan M. Taylor, “Productivity, Tradability, and the Long-Run Price Puzzle,” NBER Working Paper 10569, National Bureau of Economic Research, June 2004. A useful discussion and graph can be found in Alan M. Taylor and Mark P. Taylor, “The Purchasing Power Parity Debate,” *Journal of Economic Perspectives*, Fall 2004, pp. 135–158.

<sup>6</sup> The present author pleads guilty to having followed this incorrect convention in the RICE models.

<sup>7</sup> Neary discusses these issues and proposes an alternative that does not require homothetic preferences (J. Peter Neary, “Rationalizing the Penn World Table: True Multilateral Indices for International Comparisons of Real Income,” *American Economic Review*, December 2004, vol. 94, no. 5, p. 1411–1428).

<sup>8</sup> The term superlative was introduced by W. Erwin Diewert in “Exact and Superlative Index Numbers,” *Journal of Econometrics*, vol. 46, May 1976, pp. 115–45.

United States National Income and Product Accounts. Another well-known example is the Törnqvist index, which is used in the U.S. Bureau of Labor Statistics chained consumer price index. Appendix B describes superlative indexes in more detail and provides the formula for the Törnqvist index.

The advantage of superlative indexes is that they provide much better approximations of real income and utilities than earlier indexes such as Laspeyres indexes; this is true particularly when the relative prices of two situations are far apart. They avoid the problems of assuming constant weights or constant prices for different goods. They have corresponding disadvantages, such as greater data requirements and non-additivity, but most practitioners believe that improved approximations offset the disadvantages.

### *2.1.1. National accounts in practice in the international arena*

There are two widely used statistical sources for the national economic accounts of different countries that rely on market prices and market exchange rates to put these in a common currency. The first is the International Financial Statistics (IFS) of the IMF. This contains national income accounts and international trade data for most countries for the period from 1948, and converts these using a given year's market exchange rate. A closely related data set is the World Bank accounts. The World Bank data are conceptually similar to the IMF data. In principle, the World Bank and the IMF real GDP estimates should be identical except for a scalar multiple for each country; the growth rates of the different variables should therefore be identical.

## *2.2. Purchasing power parity (PPP) accounts*

As discussed in Section 1, it has been recognized for many years that using market exchange rates to compare the incomes, expenditures, and outputs of different countries can produce misleading estimates. By contrast, PPP accounts are based on explicit attempts to measure prices and outputs in different countries as the building blocks for estimating comparative incomes and expenditures.

### *2.2.1. Bilateral comparisons*

The basic goal of PPP accounts is to measure “true” relative incomes, outputs, or expenditures of different countries. PPP output measures the real values of output or expenditure in different countries using a standardized measuring rod. Construction of PPP output measures starts at the basic commodity level by estimating the level of real output or expenditure at a disaggregated level (shoes, women's shoes, compact automobiles, and so forth). It then calculates the relative output or real expenditure of each country as the ratio of total real value of outputs or expenditures at a common world price level to a reference country.

The simplest example would be the bilateral comparison of two countries, where the relative outputs would be compared using an index such as Laspeyres or Paasche, or a superlative index such as the Fisher or Törnqvist–Theil index. The latter index is also illustrated in Appendix B.

### *2.2.2. Multilateral comparisons*

When moving from bilateral to multilateral comparisons, major complications arise. The major difficulty is that there is no obvious and simple way to move from bilateral to multilateral measures. Whereas in time series it is natural to chain two adjacent years together, there is no obvious way to array countries so that they are next to each other like strings on a bead. Should countries be arrayed by geography, per capita incomes, price similarity, size, linguistic affinity, or what?

There are several leading contenders for the best multilateral method of constructing PPPs for a multilateral index of countries in a benchmark year. The multilateral analogue of the Törnqvist–

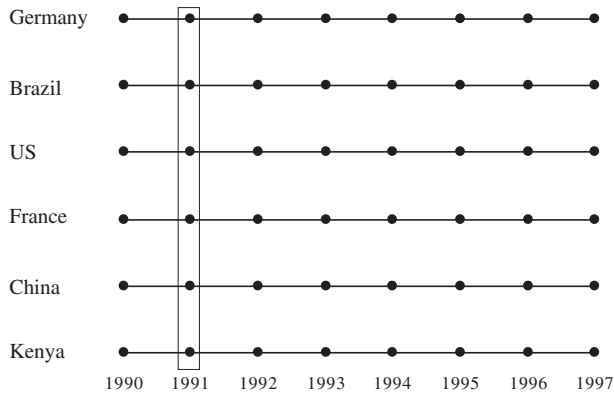


Fig. 3. Combination of time series and cross-section. The most common procedure for combining time series and cross-sections is to take a PPP cross-section for a given year and then use national-accounts growth rates to extrapolate country data forward and backwards in time. In this example, the cross-section PPP data are the vertical box for 1991, and national data are the horizontal lines from 1990 to 1997. Note that the 1997 extrapolated cross-section will not necessarily correspond to a PPP cross-section for that year. (This figure is inspired by Robert Hill, “Constructing Price Indexes Across Space and Time: The Case of the European Union,” *American Economic Review*, December 2004, vol. 94, no. 5, pp. 1379–1410.)

Theil index is known as the CCD method.<sup>9</sup> When the Fisher formula is used rather than the Törnqvist formula, it is known as the EKS approach, which is the preferred approach of the OECD and Eurostat.<sup>10</sup> Both methods construct multilateral indexes as the unweighted geometrical averages of the bilateral indexes. The well-known Penn World Tables use a technique (the Geary–Khamis formula), which estimates a set of “world prices” and then values national levels of expenditure quantities using world prices. Yet a fourth method is Hill’s use of “spanning trees,” which constructs pathways among countries based on price similarities.<sup>11</sup>

Reviews of the alternatives suggest that none of the multilateral approaches is completely satisfactory.<sup>12</sup> The unweighted formulas, such as CCD or EKS, are unbalanced because they weight Singapore as much as the United States. By contrast, the Geary–Khamis approach of the Penn World Table weights large countries more heavily, making that index similar to a Paasche index that overestimates the relative incomes of low-income countries (see the Numerical example in the next section). This area is the subject of intense research, and best practice may well change in the next few years.

### 2.2.3. Constructing time series

A final issue concerns how to construct *panels of countries* with PPP outputs or expenditures (i.e., how to combine the time series and the cross-sections). The most common practice is to combine the

<sup>9</sup> Douglas R. Caves, Laurits R. Christensen, and W. Erwin Diewert, “Multilateral Comparisons of Output, Input and Productivity using Superlative Index Numbers,” *The Economic Journal*, March 1982, pp. 73–86 (hence “CCD”).

<sup>10</sup> This is named after Eltető, Köves, and Szulc (hence “EKS”). The standard citations are O. Eltető and P. Köves, “On an index number Computation Problem in International Comparison,” (in Hungarian), *Statistikai Szemle*, vol. 42, pp. 507–518 and B. Szulc, “Index Numbers of Multilateral Regional Comparisons” (in Polish), *Przegląd Statystyczny*, vol. 3, 1964, pp. 239–254.

<sup>11</sup> Robert Hill, “Linking Countries And Regions Using Chaining Methods And Spanning Trees,” Joint World Bank — OECD Seminar On Purchasing Power Parities, January 1, February 2, 2001.

<sup>12</sup> A useful reference is Robert J. Hill, “A Taxonomy of Multilateral Methods for Making International Comparisons of Prices and Quantities,” *Review of Income and Wealth*, March 1997, 43(1), 49–69 and W. Erwin Diewert, “Axiomatic and Economic Approaches to International Comparisons,” in Alan W. Heston and Robert E. Lipsey eds., *International and Interarea Comparisons of Income, Output and Prices*, NBER Studies in Income and Wealth, vol. 61, Chicago: University of Chicago Press, 1999, pp. 13–87.

Table 1  
Comparison of per capita incomes of major OECD countries using alternative benchmark dates, data, and techniques

	Alternative PPPs for 1999, national currency/USD					Alternative estimates of GDP per capita, United States=100				
	Official 1999 PPP, based on the 1996 EKS index	Based on the 1993 EKS index	Based on the 1990 EKS index	Based on the 1985 Fisher index	Based on the 1980 Fisher index	Official estimate, based on the 1996 EKS index	Based on the 1993 EKS index	Based on the 1990 EKS index	Based on the 1985 Fisher index	Based on the 1980 Fisher index
Australia	1.30	1.30	1.30	1.41	..	75.6	75.5	75.8	69.4	..
Austria	13.6	13.7	14.2	17.0	14.8	72.8	72.2	69.6	58.2	66.7
Belgium	37.1	36.5	39.2	44.8	37.5	73.4	74.6	69.4	60.8	72.6
Canada	1.17	1.21	1.22	1.19	1.12	79.3	76.5	76.0	77.9	82.6
Czech Republic	13.4	..	..	..	..	39.5	..	..	..	..
Denmark	8.54	8.92	9.41	10.81	9.15	79.1	75.7	71.8	62.5	73.8
Finland	6.15	6.14	6.22	7.05	..	67.2	67.3	66.4	58.6	..
France <sup>a</sup>	6.63	6.34	6.32	7.23	6.22	65.2	68.2	68.5	59.8	69.6
Germany	1.98	2.05	2.12	2.52	2.18	70.4	68.1	65.7	55.3	64.0
Greece	239.2	250.3	277.6	294.2	278.5	44.7	42.8	38.5	36.4	38.4
Hungary	100.7	..	..	..	..	33.3	..	..	..	..
Iceland	85.5	89.8	..	..	..	77.8	74.1	..	..	..
Ireland	0.73	0.72	0.77	0.84	0.72	75.1	75.4	70.6	64.8	75.3
Italy	1601.1	1681.2	1675.2	1911.2	1694.7	68.2	64.9	65.1	57.1	64.4
Japan	158.7	161.6	165.9	190.0	168.3	75.3	74.0	72.1	62.9	71.0
Korea	642.9	677.7	..	..	..	47.5	45.0	..	..	..
Luxembourg	40.6	40.7	..	..	..	122.2	121.9	..	..	..
Mexico	5.61	..	..	..	..	25.0	..	..	..	..
Netherlands	1.99	2.13	2.14	2.32	2.11	77.5	72.3	72.0	66.3	72.9
New Zealand	1.46	1.46	1.52	1.84	..	55.1	54.8	52.9	43.6	..
Norway	9.50	9.38	9.76	9.71	8.16	83.1	84.3	80.9	81.4	96.9
Poland	1.84	..	..	..	..	25.6	..	..	..	..
Portugal	128.9	..	..	..	..	48.6	..	..	..	..
Slovak Republic	13.8	..	..	..	..	32.4	..	..	..	..
Spain	130.5	127.5	131.1	146.1	130.7	53.8	55.1	53.6	48.1	53.8
Sweden	9.78	9.77	9.56	10.48	..	68.0	68.1	69.6	63.5	..
Switzerland <sup>b</sup>	1.90	1.98	2.11	..	..	84.7	81.1	76.2	..	..
Turkey <sup>b</sup>	185 567	183 857	183 799	177 331	..	18.7	18.9	18.9	19.6	..
United Kingdom	0.66	0.67	0.66	0.74	0.69	67.6	66.3	66.9	60.0	63.8
United States	1.00	1.00	1.00	1.00	1.00	100.0	100.0	100.00	100.00	100.00

Source: Paul Schreyer and Dirk Pilat, "Measuring Productivity," *OECD Economic Studies*, No. 33, 2001/II, pp. 127–180.

<sup>a</sup> Includes overseas department.

<sup>b</sup> Countries still using the 1968 SNA, i.e. GDP, may be underestimated compared with other OECD countries.

cross-section of a base year (which is a PPP measure) with extrapolations forward and backward for each country using that country's prices and outputs. Fig. 3 shows schematically how the standard approach works. This procedure is used by the World Bank, the OECD, and Angus Maddison.

The time-series extrapolations in the Penn World Table follow a different procedure from most other PPP time series. They use national time series for the major components of GDP (for consumption, investment, government purchases, exports, and imports), but use weights based on world prices rather than national prices to weight the components. (This is the case for both the PWT Laspeyres and the PWT chain indexes.) This implies that the time-series growth rates in the PWT will differ from those in other PPP calculations or from MER growth rates.<sup>13</sup>

The standard procedure of extrapolating PPP cross-sections with national-accounts time series generally leads to an inconsistency between the extrapolations and the benchmarks outside the base year. How important is the inconsistency? Table 1, drawn from Schreyer and Pilat, presents results from the OECD countries. There are clear differences because of differences in the base year, different techniques, and inconsistent price data. For example, New Zealand shows a 20% difference from 1985 to 1990 (base year, data, and technique changes), while Greece shows a 10% change from 1990 to 1993 (base year and data changes).

How should we interpret the inconsistency between PPP cross-sections and national time series? I suspect that differences arise primarily because of data differences. Under ideal circumstances,<sup>14</sup> the time series and cross-sections will yield consistent estimates.

Fig. 3 (inspired by Robert Hill) is a good way of understanding the usual approach and the one recommended here. The levels of GDP for different countries are estimated using a PPP comparison for 1991 (shown as the square running vertically in 1991). The panel is then completed by using national data on real GDP for each country forward and backward in time. Under ideal circumstances, these would be consistent. In other words, we would get the same growth rate for France whether (a) we go horizontally from 1991 to 1997 using the French national data or (b) we go vertically with PPP from France to Germany in 1991, then horizontally with German national data from 1991 to 1997, then vertically back to France in 1997 using another PPP estimate. In other words, under ideal circumstances, the superlative PPP technique will yield cross-sectional and time-series consistency. The examples of New Zealand and Greece just cited, or the examples in Table 1, indicate that the "ideal circumstances" do not hold in the reality of actual data, techniques, and preferences.

### 2.3. A numerical example

A numerical example may help illustrate some of the concepts here. Table 2A shows the prices and quantities of two situations. These could be either rich and poor country, or early and late periods. For simplicity, I assume that the underlying utility functions are identical and homothetic and that the level of utility (and "real output") in the rich/late situation is ten times that in the early/poor situation. The prices of tradable goods are equal in the two situations, but the price of the non-tradables is much lower in the poor/early period. This example assumes an identical constant elasticity of substitution utility function in the two countries/periods with elasticity between the two goods of 0.67. The poor/early situation has a non-tradable price equal to one-fifth of the rich/late situation, which induces relatively high consumption of non-tradables in the poor/early situation.

<sup>13</sup> Earlier versions of PWT also changed time series on national growth rates to fit alternative cross-sections, but that practice has been discontinued in the most recent versions. Note also that, because the PWT country growth rates differ from those calculated at national prices and weights, they do not meet the criterion for "superlative PPP accounts" that is suggested here.

<sup>14</sup> By "ideal circumstance," I mean that countries have identical homothetic social indifference curves that generate the price and output observations, that data are accurate, and that time series and cross-section use the same superlative indexes.

Table 2A  
Illustrative prices and per capita quantities of two situations

Period /country	Prices		Quantities	
	Tradable	Non-tradable	Tradable	Non-tradable
Rich/late period	100.0	100.0	100.0	100.0
Poor/early period	100.0	20.0	6.3	18.4

This example shows a typical case where a rich country (or a country after development) has relatively high prices of non-tradable goods. The high prices in the rich country induce low consumption of non-tradables relative to a poor country (or a country before development).

Table 2B  
Calculation of relative outputs using different index-number formulas

Index of relative outputs (ratio of rich/late or poor/early)		
Technique	Difference from true [log, percent]	Ratio of levels [rich to poor]
MER	69.8%	20.10
Laspeyres	18.7%	12.06
Utility (true)	0.0%	10.00
Törnqvist	-0.1%	9.99
Fisher	-1.1%	9.89
Geary–Khamis	-16.8%	8.45
Paasche	-20.9%	8.12

This table shows calculations for different index number techniques for the data shown in Table 2A. The utility functions are assumed to be identical, where the rich/late country has utility or real output ten times than of the poor/early country. This example shows the typical bias of different techniques that are found in cross-section estimates.

Using these data, we can apply standard index-number formulas to obtain the relative outputs of the two situations. The results of the major techniques are shown in Table 2B. The actual utility or “real output” difference is a factor of ten by assumption. The two superlative indexes, Fisher and Törnqvist, give very close approximations, which shows why statisticians like these indexes. The Paasche index (reflecting rich/late weights) shows the expected result that the rich/late situation has relatively low output (or that the growth rate is relatively low), while the Laspeyres index indicates the opposite. This example also shows one of the problems with the standard Geary–Khamis (GK) technique, which is that it resembles a Paasche index because it puts heavier weights on large countries; in PPP calculations, GK tends to overestimate the relative incomes of poor.

Finally, the MER calculation provides a highly exaggerated estimate of the difference, indicating that the relative output level of the rich/late situation is much higher than reality. MER is also way outside the Paasche–Laspeyres bound, which shows the general proposition that MER is simply an incorrect technique for calculating relative outputs.

This numerical example also shows a fundamental and subtle point about the superlative PPP technique: *combining the PPP levels and the national growth rates using superlative indexes is in fact an internally consistent technique for calculating outputs across time and space.* This paradoxical result relies on the fact that ideal spatial techniques share the same conceptual basis as ideal intertemporal techniques.

#### 2.4. Comparison of growth rates in different accounts

The empirical implementation of PPP accounting systems has made great progress in the last three decades. At present, the most comprehensive and widely used set of PPP accounts is the “Penn World

Table,” or PWT. The current version, PWT 6.1, is the sixth round of comparisons and relies largely on international price data collected for 1996. PWT 6.1 provides PPP and national income accounts converted at world prices for 168 countries covering the period 1950 to 2000.<sup>15</sup> This approach estimates the base year national income by calculating the value of domestic expenditures at world prices. World prices are essentially the average of different countries’ prices weighted by country expenditures. To construct time series for each country, PWT 6.1 projects the base-year GDP over time using data on the components from the national accounts of individual countries, as discussed above.

There are other PPP series available using different approaches. Maddison extends his PPP estimates for a smaller set of countries much further back in time (in some cases back to 1820). His approach is similar to that of the World Bank, using national time series to project individual countries over time. The OECD and Eurostat have made comparisons of a limited set of countries (primarily high-income countries of the OECD and Europe, respectively). These results also use the benchmark for a single year and national accounts estimates to project over time.

How do the different estimates of growth compare? In principle, if they use national accounts extrapolators and are based on the same underlying data (that is, have not been revised), the growth rates of different countries should be identical. Studies indicate that in fact there are differences in estimated growth rates among the different countries between PPP and MER accounting systems.<sup>16</sup> Appendix A shows a compilation for five important countries for six different data sets. The first three sets of bars are growth rates from the World Bank data set (in constant U.S. dollars, in constant local currency, and in constant PPP). With a few exceptions, these are identical, which indicates that the PPP is indeed constructed from the national real GDP series.

The last bar is the IMF series, which uses MER and national growth rates. These also are consistent with the World Bank series (which are drawn from the IMF), although there are a few inconsistencies, which may be data revisions.

The fourth and fifth bars are the PPP estimates from Maddison and PWT, respectively. Small differences among the series may arise because the PPP estimates from PWT and Maddison are based on earlier national data (from around 2000) and differ from the World Bank growth rates because of data revisions. Looking at high-income countries, the PPP estimates are quite consistent with the MER estimates for Japan and Mexico. They show small differences with India and the U.S. However, they are very different for China because China received special handling in most PPP accounts. The reasons for the remaining differences between the PWT and other PPP results are not obvious, but they probably relate to the differences in weights, as discussed above. Users of the PWT data should be warned that there are differences in concepts and growth rates among different PPP data sources.

### 3. Data quality in PPP and national accounting systems

#### 3.1. *Comprehensiveness of accounting systems*

How comprehensive are actual PPP accounting systems as compared to national economic accounts? Current best practice in national economic accounts strives to build a set of integrated accounts. Construction of national accounts is currently coordinated through the internationally

<sup>15</sup> The description is from [http://pwt.econ.upenn.edu/php\\_site/pwt\\_index.php](http://pwt.econ.upenn.edu/php_site/pwt_index.php).

<sup>16</sup> See Daniel A. Nuxoll, “Differences in Relative Prices and International Differences in Growth Rates,” *American Economic Review*, vol. 84, no. 5, Dec. 1994, pp. 1423–1436. He found that the growth rate in the PWT was slightly lower than that in domestic prices for poor countries relative to rich countries, which is the opposite of what was expected.

developed system of national accounts, or SNA, the latest being 1993.<sup>17</sup> The major feature of the SNA is that it contains multiple sets of accounts linking production, expenditure, income, consumption, accumulation, and wealth. Few national accounts have achieved complete integration, but this is a stated goal of the United States and many other countries. At present, most high-income countries have income and expenditure accounts, and some progress has been made on developing production and capital accounts. All national accounts are based on market prices.

By contrast, virtually all work on PPP accounting has been limited to expenditure accounts. For technical reasons, deriving production accounts for industries (agriculture, energy, manufacturing, services, etc.) has proven much more difficult for both the national accounts and for international comparisons. The major difficulties in developing production accounts are the need to develop comprehensive data on intermediate goods and services, as well as capital goods and capital services, and the requirement to develop double-deflation techniques to estimate real value added in production by sector. Because of these difficulties, work on developing PPP production accounts is still in its infancy, or in some cases awaiting conception.<sup>18</sup> The lack of production accounts is a major practical limitation for models, such as for energy sectors, that have an important component in production.

### 3.2. *Quality, matching, and comprehensiveness of data*

A further issue is the matching, comprehensiveness, and quality of the data underlying PPP accounts. One difficulty is that the bundle of goods differs fundamentally across countries. For example, alcohol is prohibited in many Muslim countries, while expenditures on alcoholic beverages are substantial in many OECD countries. Some of the fancy medical procedures in high-income countries will not be found in poor countries. Hence, there will be a major problem of matching goods and services of different countries.<sup>19</sup>

Further, there is no question that the size of the budgets and the quality of the data going to prepare national accounts are orders of magnitude greater than going into preparing PPP data and accounts. Some examples will suggest the differences. For the 1993 study, the Netherlands contributed 15 prices for men's footwear in the International Comparison of Prices (ICP), whereas in a recent year it collected 7358 observations on sport shoes. The Netherlands contributed 110 prices for furniture and fixtures to the ICP, while collecting 3501 prices for light bulbs. Finland collects about 1 million price observations annually for its consumer price index, while there are 2500 items included in the consumption bundle for the OECD region in its international price comparison.<sup>20</sup> The U.S. Bureau of

<sup>17</sup> The SNA, or System of National Accounts, developed under the aegis of the United Nations and other international agencies, is a set of concepts, definitions, classifications, and accounting rules. The latest SNA is from 1993 and can be found at <http://esa.un.org/unsd/sna1993/introduction.asp>. The United States has adopted most of the principles in the SNA.

<sup>18</sup> One of the major efforts in this area is from the International Comparison of Output and Productivity (ICOP) project. See Bart van Ark and Marcel Timmer, "PPP and International Productivity Comparisons: Bottlenecks and New Directions," Groningen Growth and Development and The Conference Board Paper for ILO Seminar On Productivity Comparisons, January 26, 2001, Geneva, and Joint World Bank—OECD Seminar On Purchasing Power Parities, January 30–February 2, 2001, Washington D.C. Also see Robert Feenstra, Alan Heston, Marcel Timmer, and Haiyan Deng, "Estimating Real Production and Expenditures Across Nations: A Proposal for Improving Existing Practice," Research Paper, Center for International Comparisons, University of Pennsylvania, Philadelphia, PA, June, 2004.

<sup>19</sup> I am indebted to Erwin Diewert for the points in this paragraph.

<sup>20</sup> Nicole Jonker, Carsten Folkertsma, and Harry Blijenberg, *An Empirical Analysis Of Price Setting Behaviour In The Netherlands In The Period 1998–2003 Using Micro Data*, European Central Bank Working Paper Series No. 413, November 2004. Jouko Vilmunen and Helinä Laakkonen, "How Often do Prices Change in Finland? Micro-Level Evidence from the CPI," Bank of Finland Working Paper, November 2004.

Economic Analysis devotes a tiny fraction of the time of one person to PPP estimates out of its professional staff of 150. As best can be determined, data collection in Europe of consumer prices for PPP programs is around \$1 million per year, as compared to a budget of around \$100 million for the U.S. CPI program.

The general view of experts in this area is that the price and output data and aggregates are much more accurate for the national accounts of most high-income countries than for the various data collections generating PPP incomes and prices.

#### **4. Correspondence of price and output measures to the economic theories underlying the models**

The next issue is which system of accounts better fits the conceptual framework of economic and environmental models. There will be no universally correct answer to this question given the differing structure of models. For this discussion, I will address four different aspects of modeling frameworks: engineering-programming, econometric, accounting, and estimation.

An engineering-programming framework is often used in modules of energy models that rely on specific engineering production functions to model inputs and outputs of the energy systems. For example, many models have physical inputs of fuels (oil, gas,  $U^{235}$ , etc.) as well as electricity output (kWh or passenger-miles). In such a framework, the costs of different inputs and outputs are linked inside the larger model using the market prices of the region or country in which the activity took place.

This approach relies upon the “duality principle” under which a competitive market has the same prices and outputs as an optimized engineering-programming model. For such a modeling approach to provide accurate answers, it is necessary that the costs, prices, and interest rates used in the models correspond to the prices that are faced in the competitive market. In this respect, then, it would be necessary to use the actual market prices in the modeling. If the models used “world prices” (in the fashion calculated by some PPP estimates) for individual inputs and outputs (say of oil or gasoline), the output generated by the model would be biased away from the competitive market prices.

A second general approach used in models is an econometric one. For example, many models contain demand systems in which energy demands are related to prices, incomes, outputs, as well as geophysical features. Most econometric estimates at a detailed level (say, of the demand for oil, coal, and natural gas that would be used to generate carbon emissions) relate demand to actual market prices. Clearly, it would be necessary to use actual market prices to run the models and generate accurate projections.

Moreover, it would make no sense to use commodity-specific “world prices” either in the estimation or in the projections. Suppose models were attempting to determine why gasoline usage was relatively high in the U.S. and relatively low in Europe. If the estimates were commodity-specific world prices, such as those generated by the Geary–Khamis technique for the World Penn Tables, then the different price levels in different countries would get lost in the estimation. Consumers respond to local prices, not to commodity-specific world prices.

This point becomes even clearer when issues of international trade are considered. As one economist put it, trade flows uphill from low-cost regions to high-price regions. Empirical models of international trade (such as those based on Heckscher–Ohlin principles) generate predictions of trade flows based on different prices. It would completely undo these insights, and make empirical trade work involving relative prices impossible, if our accounting systems were to move to uniform world prices instead of actual market prices.

Similarly, models sometimes project the implications of various combinations of energy or climate-change policies. These might consider policies such as the Kyoto Protocol, where participating countries have high carbon taxes (or their equivalent) and non-participants have zero

carbon taxes. It would clearly make no sense to use uniform world prices in such calculations — the models need to have differentiated responses to different prices.

The final point is that most modeling systems involve accounting identities (such as income equals output or saving equals investment). It is clearly important that the accounting systems be compatible. This would require that the income and output identities be consistent with the engineering and econometric accounts.

Weighing these considerations, the analytical framework of most economic models, especially models involving energy and international trade, involves relationships between outputs and actual market prices. The use of world prices will provide distorted economic estimates and cannot serve as a reliable set of price and output data to use for projections. It is essential to build accurate models and make useful projections that the models use actual market prices faced by market participants.

## 5. Accounting for convergence

One of the major questions concerning future growth rates involves whether countries will converge to a common level of per capita income or output, or perhaps converge partially over some period. How much convergence has taken place or is taking place is by itself a controversial issue. However, let us put aside the empirical questions about historical and future convergence to ask which accounting system would best estimate growth rates should convergence occur.

Let us use India as an example. How much would India need to grow to reach U.S. levels of output per capita? The World Bank's estimate of the ratio of per capita GDP in India to that in the U.S. for 2003 was 7.7 using the PPP measure and 1.4% using the MER measure. If a modeler wanted to assume that India converged to U.S. outputs and prices over the next half-century, what should be done? Convergence would require India's 50-year logarithmic growth rate exceed the U.S. rate by 5.1% per year using the PPP measure and by 8.5% per year using the MER measure. Which is the correct approach?

The answer is simple: In principle, the superlative PPP method would provide the correct answer. In other words, a modeler should start with the best estimate of the true current relative real outputs (i.e., the best PPP number) and then use correctly measured national growth rates at national prices. This is the correct approach because the measure of the output difference between the U.S. and India in 2003 is calculated in the same way as the measure of the output difference between India in 2003 and 2053. In our convergence calculation, by assumption, India in 2053 has the same economic structure as the U.S. in 2003. Therefore, in principle, using the same index-number formula, the growth factor over time would equal the growth factor over space.

This discussion is heavily larded with “in principle.” This is to remind us that the index-number calculations across space and time should be based on similar techniques (e.g., superlative indexes such as Törnqvist or Fisher across both space and time). Moreover, and more significant for this example, we might worry about Indian data quality as well as the consistency between Indian data and the data on which international comparisons are made. Notwithstanding these reservations, it is likely that the PPP imprecision is small relative to the MER bias. If the PPP numbers were imprecise by a factor of plus or minus 20%,<sup>21</sup> then the

<sup>21</sup> The estimates of precision for India in Kravis, Heston, and Summers, *op. cit.*, Chapter 3 are between 10% and 20% for a 95% confidence interval based on differences in data and techniques.

fifty-year convergence growth rates would lie in the range of 4.8 to 5.5% per year. This range of imprecision pales in comparison to the difference between these numbers and the MER estimate of 8.5% per year.

Does this analysis indicate whether estimates of future emissions are biased because they relied upon MER rather than PPP estimates of output? I address that issue further in the next section, but from an analytical vantage point that question cannot be answered. To answer that question would require looking not only at the GDP growth rates but also at the estimated changes in energy- or carbon-intensities. That question has not been addressed here. Rather, this discussion is limited to the question of the best practice for measuring output both in general and in convergence exercises.

## **6. Which system of accounts provides more accurate projections of future economic, energy, and emissions trends?**

Most of the heat in the debate about PPP versus MER has focused on which accounting system is better for constructing scenarios in the context of projecting future economic activity, energy use, emissions patterns, and climate change. Most groups who have constructed global models incorporating economic modules of production, consumption, and energy production have relied upon actual market prices and MER to calibrate their models and make their forecasts. What is the best system to use given the necessary compromises between data quality, model tractability, and theoretical fit? There are several points that are addressed in this section.

### *6.1. Does it make any difference which of the two PPP systems is used?*

It definitely makes a major practical difference whether modelers use the world-price PPP accounts or the superlative PPP accounts (see Fig. 1 above). The superlative PPP accounts are very close to the existing approaches and require relatively few technical conversion issues; using the superlative PPP approach requires adjusting levels of income and output but not growth rates or relative prices. Constructing the world-price PPP accounts is an enormous, probably unachievable, and analytically unattractive approach; it would require constructing an entirely new set of income, capital, interest rate, and price accounts over space and time.

### *6.2. Is there a one-to-one mapping from MER to PPP?*

If there is a one-to-one linear mapping from MER to PPP accounts, then it might be possible to convert from MER to PPP relatively easily. In the simplest case, the difference would be similar to converting from the English to the metric system of measurement. However, a linear conversion is unlikely to hold in practice because of non-linearities in the aggregation of countries and in behavioral relationships.

### *6.3. Are there problems with reduced form and behavioral relationships?*

Some of the difficulty in constructing global energy-environment-economy models arises because they employ multiple behavioral relationships as well as parameterized or “reduced-form” relationships between different aggregate variables. For example, the rate of “decarbonization” and the rate of “autonomous energy efficiency improvement” (AEEI) are rules of thumb derived from historical relationships between output growth and energy use or carbon emissions. Similarly, price elasticities of demand are calculated using actual market prices and incomes. Depending upon how

these relationships were estimated, they might need to be re-estimated and recalibrated if the accounting system were to change to the superlative PPP approach. But even here, attention to specifications is necessary. If the estimates were based on panel data methods with country fixed effects, then the estimates would continue to apply to the superlative PPP system; however, if the estimates included cross-sectional income elasticities, then the relationships would need to be re-estimated using the PPP incomes.

It is no small task to review all these relationships and to recalibrate models using PPP income and output data. I might mention that a preliminary version of the RICE-2000 model used PPP output levels rather than MER output levels. This was abandoned, and the modeling effort continued with MER, because it was thought to be too difficult to make all the necessary adjustments in a time-and-space-consistent regionalized energy and economic model.

#### *6.4. Will disaggregated models solve the problem?*

An alternative approach for constructing global models would be highly disaggregated “general equilibrium” models that disaggregate all sectors, incorporate multiple sectors of production and consumption, and use actual national prices, outputs, and incomes. In this fantasy world of the future, the MER versus PPP problem would disappear because, in effect, all processes of the actual economy would be reproduced accurately. If the actual heterogeneity of outputs were captured in detailed models, then income and output indexes can be calculated that reflect both PPP and MER measures. In other words, this extraordinary model would include the correct actual prices and quantities of all traded and non-traded goods. While this approach will solve the problem in principle, it is unlikely that price and quantity data will be available for all sectors and most important countries, particularly developing countries, in the foreseeable future.

#### *6.5. Does moving from MER to PPP change the projections?*

Does it make any difference in practice which set of accounts is used for the projections? Some preliminary runs suggest it might. McKibbin, Pearce, and Stegman find that, when they recalibrate their model using PPP growth rates with their convergence conventions, they estimate MER emissions to be 40% higher than PPP emissions in 2100.<sup>22</sup> Using the MERGE model, Manne and Richels calculate that MER exchange rates give 17% higher global emissions than PPP rates in 2100.<sup>23</sup> Tol reports that alternative exchange rate approaches make substantial differences to emissions in the FUND model, with a range of about 40% in 2100 between different exchange rate and economic assumptions.<sup>24</sup> Early versions of the RICE-2000 model calibrated (probably incorrectly) to PPP output levels also showed substantially different projections than the later ones using MER output levels.

By contrast, some modelers who have corrected both growth rates and energy intensities report little difference in emissions. For example, Alfsen and Holtmark use a small economic model to estimate emissions and find essentially no differences between PPP and MER projections of

<sup>22</sup> Warwick J. McKibbin, David Pearce, and Alison Stegman, “Can the IPCC SRES be Improved?” Manuscript, May 2004.

<sup>23</sup> Alan Manne and Richard Richels, “Market Exchange Rates or Purchasing Power Parity: Does the Choice Make a Difference to the Climate Debate?” Manuscript, August 20, 2003, downloaded from <http://www.stanford.edu/group/MERGE/biblio.htm>.

<sup>24</sup> Richard S.J. Tol, “Exchange Rates and Climate Change: An Application of FUND,” January 16, 2005, forthcoming in *Climate Change*.

emissions.<sup>25</sup> Another set of model estimates, by Nakićenović and his colleagues, adjusted the output levels and growth rates along with the energy intensities. For these runs, the doubly corrected emissions projections differ very little between the PPP and the MER calibrations.<sup>26</sup>

At this point, the jury is out on how much using PPP as compared to MER will affect aggregate emissions. Some of the emissions differences reported above probably arose because the authors did not recalibrate all the behavioral and reduced-form relationships in their models.

The uncertainties about MER versus PPP should be put in perspective. This is but one of the many uncertainties that modelers face in constructing their models. It is by no means clear that exchange-rate-conversion issues are as important as uncertainties about population, technological change, or the many geophysical uncertainties. More work needs to be done, but it would be a mistake to concentrate all efforts on correcting exchange-rate issues at the expense of other uncertainties.

#### 6.6. Does use of MER bias econometric estimates?

Using MER output levels to estimate behavioral relationships or reduced-form equations may lead to biased estimates. The reason is that there are errors in measurement of income when using MER, and these can produce incorrect results. These points can be illustrated by Fig. 4. This shows the trajectory of emissions (on the vertical axis) and “output” (on the horizontal axis). The solid heavy PPP line shows China and the U.S. in periods 1 and 2. This example assumes that the elasticity of emissions with respect to (correctly measured) output is constant and that emissions intensities are equal at equal outputs. As the U.S. and China grow, they move up the PPP line as shown.

The emissions-output lines using MER are more complicated. The solid line for period 1 shows how the data for China and the U.S. are joined, with the implicit elasticity for MER lower than that for PPP because the output differential is higher. The problem is that when we look at the cross-section for the next period, the line has shifted. Moreover, if relative prices have converged, as we would expect under the Harrod–Balassa–Samuelson effect, then the slope of the cross-section MER line tilts toward that of the (true) cross-section of the PPP line, as shown in the figure.

Using this figure, we can see that if we estimate emissions-output elasticities from the MER line, we get a biased estimate. For example, if we project second period emissions from the cross-section for the first period, both the U.S. and China will be off the estimated relationship. By contrast, if we project future emissions from the PPP line, the second period projections will be on the correct cross-section. The difficulty arises because the MER cross-section gives a biased and time-varying relationship between emissions and output. Note that while this example uses emissions, the same problem would apply to other variables, such as energy demand.

#### 6.7. How should we aggregate countries using the superlative PPP system?

In global economic-environmental models, variables are aggregated up from smaller units to larger aggregates. For example, modelers will always need to aggregate subnational regions (such as states of the United States) and will usually need to aggregate countries (such as the various nations within the European Union). What is the most accurate approach for aggregation? How should such aggregation take place when using the superlative PPP system?

<sup>25</sup> Bjart J. Holtmark and Knut H. Alfsen, “PPP-correction of the IPCC emission scenarios — does it matter?” Statistics Norway, Research Department Discussion Papers, No. 366, February 2004.

<sup>26</sup> Nebojša Nakićenović, personal communications, December 19, 2004 and January 12, 2005. Their results are reported in N. Nakićenović, A. Grübler and A. McDonald, Eds., *Global Energy Perspectives*, Cambridge University Press, Cambridge, UK, 1998.

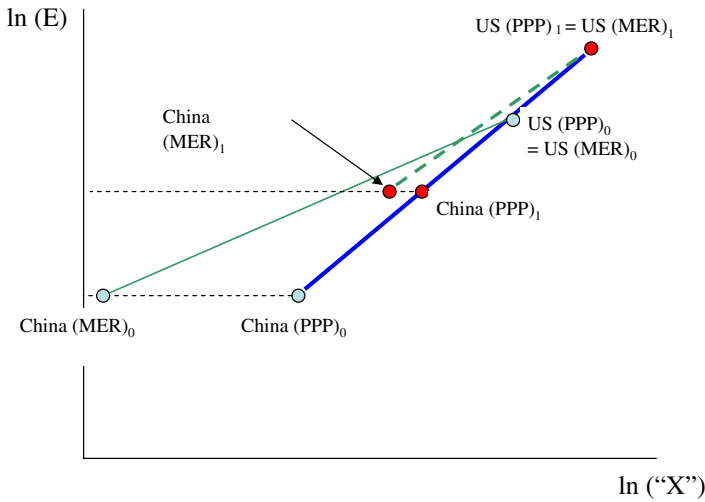


Fig. 4. Illustration of cross-sections of MER and PPP. This figure illustrates the potential for biased estimates when MER outputs are used. The true relationship lies along with heavy solid PPP line. If MER cross-sectional relationships are used, a first-period cross-section is given by the flatter thin solid line between  $\text{China (MER)}_0$  and  $\text{US (MER)}_0$ . Using that cross-section to project future points will provide biased estimates that lie along the original cross-section but are inconsistent with future data. Using the (correct) PPP outputs will provide an accurate projection of future emissions.

This turns out to be a non-trivial question. I investigate this issue in Appendix C of this study. That analysis decomposes aggregate energy or carbon intensity under different kinds of indexes and derives an equation for the determinants of the aggregate intensity. It concludes that the best approach for constructing energy intensities or carbon intensities is to construct an aggregate growth index by weighting the individual regional growth rates using superlative PPP output weights.

There are two reasons for this finding, one theoretical and one empirical. The theoretical reason is that the major term in the equation weights individual-region growth rates in carbon intensity or energy intensity using nominal output weight. Since the most accurate nominal output weights are the superlative PPP weights, these should be used in aggregation. They will produce the most consistent series over time. The second reason, and one subject to further inquiry, is that the equation contains an aggregation error that depends upon the difference in energy or carbon intensities across regions that are aggregated. This error term is likely to be smaller when superlative PPP weights are used. The reason is that the differences in energy-output ratios and carbon-output ratios have been found to be smaller when PPP outputs are used than when MER outputs are used. Hence, the error that will arise from projections based on aggregating individual region or country intensities will be smaller when superlative PPP weights are used since the error term will be smaller.

## 7. Summary

The major conclusions of this study are as follows. First, current measures of relative outputs and incomes using market exchange rates (MER) provide inaccurate estimates of relative incomes and outputs. Moreover, the errors using MER in today's world are large — underestimating incomes in low-income countries by a factor of at least three relative to high-income countries. Different techniques for measuring real incomes and outputs using PPP valuations give different estimates,

but they all indicate that MER valuations are highly misleading. While we might be reluctant to employ PPP estimates, in this area it is better to be imprecisely right than precisely wrong.

Second, purchasing power parity (PPP) accounts are designed to provide accurate estimates of “real” incomes, outputs, and expenditures in countries with widely varying price structures. Current practice in construction of PPP accounts is to use multilateral indexes that estimate the cost in a benchmark year of purchasing a nation’s quantity bundle in other countries or at world prices. In most PPP accounting systems, these estimates for benchmark years are extrapolated to non-benchmark years using growth rates from national accounts using national prices and quantities. Under ideal circumstances, the *superlative PPP technique*, using PPP incomes for estimating cross-sections and national prices and weights for calculating growth rates, will provide accurate and consistent cross-sections and time series of real income and output.

Third, PPP accounts today are conceptually unsettled and empirically imprecise relative to construction of national income and product accounts and national price indexes. PPP accounts are limited to expenditures in a sample of countries. There are no comprehensive PPP production accounts. Moreover, the databases used for constructing PPP accounts are perhaps two orders of magnitude more sparse than those of national economic accounts and price indexes. The sparse nature of PPP accounts suggests that the cross-sectional output estimates might be more accurate if constructed from an average of years rather than a single benchmark year; additionally, time-series estimates are most reliably drawn from national income accounts data than from PPP data.

Fourth, for purposes of constructing detailed global models of energy and the environment, substituting world prices (individual-commodity PPP valuations) for actual local or national prices is not recommended. A major problem with this approach is that national behavior is conditioned by actual prices, not by an artificial set of uniform world prices. A particularly important example is the response of different countries to nationally different global warming policies; using uniform world prices would defeat attempts to measure differential national responses to policies (such as participation or non-participation in the Kyoto Protocol).

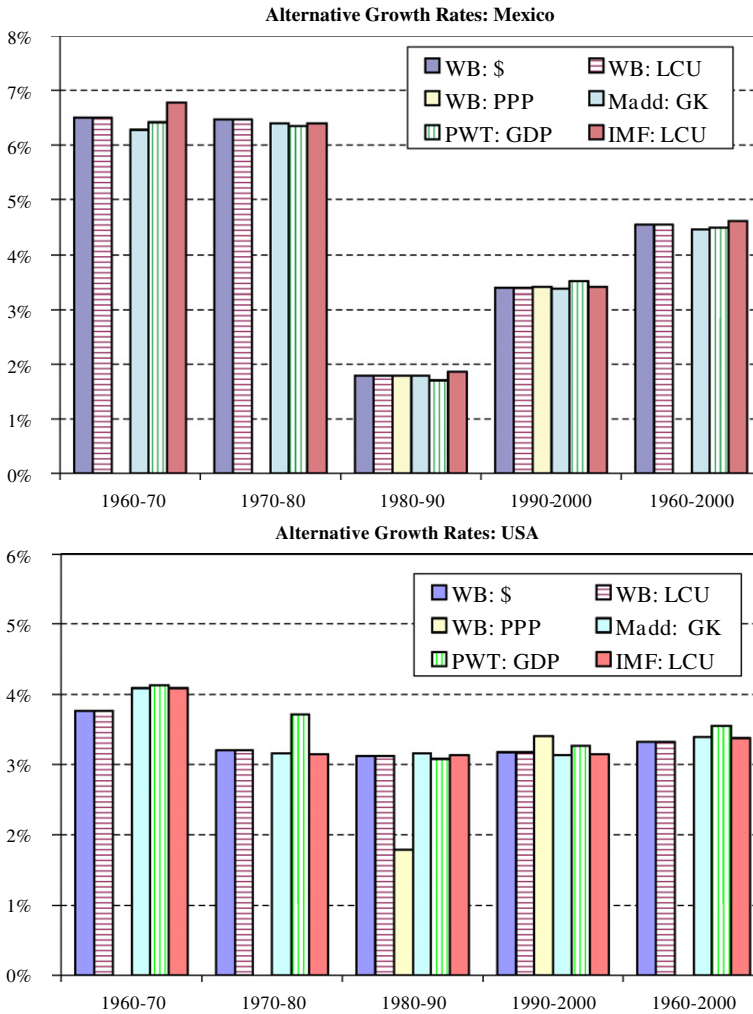
Fifth, when calculating convergence among different countries, modelers should consider the superlative PPP technique described here. That is, convergence should use true (PPP) measures of output differentials and growth rates at national prices. These will provide accurate estimates of the growth factor required for countries to attain the same output and price structure. In making these projections, it would be best to use consistent superlative index calculations and data for the cross-sections and time series.

Finally, results to date indicate that substituting PPP for MER will make a difference in projecting emissions, but it is not clear that models have been correctly calibrated. PPP incomes are likely to be better weights in constructing aggregates of variables. In moving to PPP cross-sectional valuations, modelers will need to consider whether behavioral and reduced-form relationships in the models have been correctly estimated. However, data willing, the superlative PPP approach will provide more accurate output estimates, will ensure that relative incomes and outputs correspond better to realistic magnitudes, and can ensure that welfare comparisons among regions are more accurate.

## Acknowledgment

The author is grateful for helpful comments from Erwin Diewert, Arnulf Grübler, Warwick McKibbin, Nebojša Nakićenović, Richard Tol, Jack Triplett, John Weyant, as well as a referee.





Note on symbols. “WB: \$” is the World Bank estimate of growth rates in MER in U.S. dollars; “WB: LCU” is the World Bank estimate of growth rates in local currency units; “WB: PPP” is the World Bank estimate of growth rates in PPP; “Madd: GK” is Maddison’s estimate of growth rates in PPP using the Geary–Khamis technique; “PWT: GDP” is the Penn World Table estimate of growth rates in PPP; “IMF: LCU” is the IMF estimate of growth rates in MER in local currency units. All estimates are for real growth in gross domestic product.

Sources: The World Bank: *The World Development Indicators on CD-ROM*, April 2004; International Monetary Fund, *International Financial Statistics*, Data Base on CD-ROM, December 2003; PWT: From website at <http://pwt.econ.upenn.edu/icp.html> as of December 2004; Angus Maddison, *The World Economy: A Millennial Perspective*, Development Center Studies, OECD, Paris, 2001; Maddison, A. (2003), *The World Economy: Historical Statistics*, Development Center Studies, OECD, Paris, 2003.

## Appendix B. Superlative indexes of output in time series and cross-section

Superlative price and output indexes are gradually replacing earlier approaches in national income and product accounts and in consumer price indexes. This appendix provides an algebraic example of superlative indexes for both time series and cross-sections.

An index number function is called *superlative* if it can provide a second order approximation to an arbitrary twice-differentiable linearly homogeneous function. There are many examples of superlative index number function, of which some of the best known are the Fisher ideal and the Törnqvist index. This appendix illustrates the index formula for the Törnqvist index, but in practice the results are very similar for the Fisher ideal index.

### *Index for time series*

The *Törnqvist index* of output growth is a geometric average of the growth rates of the different goods and services where the weights are the average nominal output shares of different goods in two periods or countries. To start with the example of growth over time, suppose there are  $n$  goods in two periods, 1 and 2. Let  $x_{i,t}$  be the output of good  $i$  in period  $t$ ;  $p_{i,t}$  be the price of good  $i$  in period  $t$ ;  $w_{i,t} = p_{i,t}x_{i,t} / \sum_{j=1}^n p_{j,t}x_{j,t}$  be the share of good  $i$  in total nominal output in period  $t$ ; and  $Z_t$  be an index of total output in period  $t$ . Then the Törnqvist index of the logarithmic growth of total output from period 1 to period 2 is  $\ln(Z_2/Z_1)$ :

$$\text{Output growth} = \ln(Z_2/Z_1) = \sum_{i=1}^n [(w_{i,1} + w_{i,2})/2] \ln(x_{i,2}/x_{i,1}) \quad (\text{B.1})$$

In other words, the aggregate logarithmic growth rate is the weighted average of the logarithmic growth rates of individual sectors, where the weights are the average nominal output shares of the different sectors. Time series of growth rates are then constructed by chaining the growth rates of different periods together.

### *Index for cross-section*

A superlative index of output differences between two countries is an exact analog of the output growth of a country between two periods. To estimate the difference in output between two countries, the *Törnqvist–Theil index* is the cross-section analog of the Törnqvist time-series index. This index is identical to the time-series index in Eq. (B.1), but the indexes are interpreted as belonging to different *countries* rather than different *periods*. Hence, the Törnqvist–Theil cross-section index calculates the difference between the outputs of countries B and A as:

$$\text{Output difference} = \ln(Z_B/Z_A) = \sum_{i=1}^n [(w_{i,A} + w_{i,B})/2] \ln(x_{i,B}/x_{i,A}) \quad (\text{B.2})$$

The notation in Eq. (B.2) is identical to that in Eq. (B.1) except we have substituted A and B (countries) for 1 and 2 (time periods). In this approach, the difference between the outputs of different countries is calculated as the geometric average of the outputs of the two countries where the weights are the average nominal output shares of the goods in the two

countries. The difference between the time series and the cross-section is that the weights are average across time periods for the time series and across countries for the cross-section.

As with the time-series version, there are the usual cross-sectional indexes of output growth such as the Fisher ideal.

### Appendix C. Aggregating energy and carbon intensities by region

Changes in energy or carbon intensities are generally built up from smaller units (either industries or countries). The question arises how these should be aggregated. I review this question both for fixed-weight output indexes and for superlative indexes.

#### *Intensity growth with fixed weights (“old-style output”)*

We begin with a review of the decomposition of intensity growth using fixed-weighted quantity indexes. Here, output is a Laspeyres index of the “old style” — that is, with fixed weights. These indexes are still widely used in many countries. Consider aggregates of output ( $X_t$ ), energy inputs ( $E_t$ ), and energy productivity ( $A_t = X_t/E_t$ ). The inverse of energy productivity is energy intensity,  $C_t = E_t/X_t$ . The aggregates are the weighted sums of industry output and energy ( $X_{i,t}$  and  $E_{i,t}$ ), where for simplicity we measure output in base-year prices so that they can be summed to form the aggregates. We can rewrite these as built up from industry values ( $i = 1, \dots, n$ ) as follows:

$$X_t = \sum_{i=1}^n X_{i,t}$$

$$E_t = \sum_{j=1}^n E_{j,t}$$

Aggregate energy productivity is:

$$A_t = X_t/E_t = \left( \sum_{i=1}^n X_{i,t} \right) / \left( \sum_{j=1}^n E_{j,t} \right) \tag{C.1}$$

For simplicity, I will begin by analyzing growth in continuous time under the assumption of smooth series; the discrete analysis is extremely messy. Let  $g(\cdot)$  be the logarithmic growth rate. Taking the logarithmic derivative of Eq. (C.1) yields:

$$g(A_t) = g(X_t) - g(E_t) = \sum_{i=1}^n (\dot{X}_{i,t}/X_{i,t})(X_{i,t}/X_t) - \sum_{i=1}^n (\dot{E}_{i,t}/E_{i,t})(E_{i,t}/E_t)$$

$$g(A_t) = \sum_{i=1}^n g(X_{i,t})z_{i,t} - \sum_{i=1}^n g(E_{i,t})w_{i,t} \tag{C.2}$$

where  $z_{i,t} = X_{i,t}/X_t$  = the share of real output of industry  $i$  in total real output and  $w_{i,t} = E_{i,t}/E_t$  = the share of energy use of industry  $i$  in total energy use. Dots over variables indicate time derivatives.

Next add and subtract  $\sum_{i=1}^n g(E_{i,t})z_{i,t}$  from the right-hand side of Eq. (C.2):

$$g(A_t) = \sum_{i=1}^n g(A_{i,t})z_{i,t} + \sum_{i=1}^n g(E_{i,t})[z_{i,t} - w_{i,t}] \tag{C.3}$$

The first term in Eq. (C.3) is the weighted average growth of energy productivity, while the second term is a shift effect. This result is very similar to that for superlative indexes, so I will discuss each of the terms in the next section.

*Intensity with superlative indexes (“new-style output”)*

Consider next the decomposition of intensity growth where output is measured using superlative indexes, focusing primarily on the Törnqvist index. This approach is particularly relevant as countries move to superlative indexes in their national income and product accounts. Again, using continuous time, we can now rewrite Eq. (C.2) as

$$g(A_t) = \sum_{i=1}^n g(X_{i,t})\sigma_{i,t} - \sum_{i=1}^n g(E_{i,t})w_{i,t} \quad (\text{C.4})$$

where  $\sigma_{i,t}$  is the share of industry  $i$  in nominal output for period  $t$ . Eq. (C.4) differs from Eq. (C.2) only in that the growth of individual sector outputs is now weighted by the nominal output shares over the period rather than “real” shares. Using continuous time for simplicity and then adding and subtracting  $\sum_{i=1}^n g(E_{i,t})\sigma_{i,t}$  yields:

$$g(A_t) = \sum_{i=1}^n g(A_{i,t})\sigma_{i,t} + \sum_{i=1}^n g(E_{i,t})[\sigma_{i,t} - w_{i,t}] \quad (\text{C.5})$$

or rewriting this as the growth of energy intensity,  $g(C_t)$ , we obtain:

$$g(C_t) = \sum_{i=1}^n g(C_{i,t})\sigma_{i,t} - \sum_{i=1}^n g(E_{i,t})[\sigma_{i,t} - w_{i,t}] \quad (\text{C.6})$$

The interpretation of Eq. (C.6) is that the growth of aggregate energy intensity is, by the first term, equal to the weighted average growth of energy intensity of different countries, where the weights are the shares in nominal output. This is essentially a Törnqvist index of the growth rates of country intensities. However, a correction factor must be added, which is equal to the average growth of energy use, weighted by the differences in the shares of nominal output and energy use by industry or country.

The important application for the present purposes is aggregation over countries. Since the term  $[\sigma_{i,t} - w_{i,t}]$  reflects the differences in energy intensities of different countries, the correction factor will be larger as energy-intensities differ over space and time.

This analysis suggests that the best approach for constructing energy intensities or carbon intensities is to use superlative or nominal PPP weights for countries. There are two reasons for this, one theoretical and one empirical. The theoretical reason is that the weights on the first term in Eq. (C.6) are the nominal output weights. Since the most accurate nominal output weights are the superlative PPP weights, these should be used in aggregation. They will produce the most consistent series over time. The second reason, and one subject to further inquiry, is that the correction factor in the second term of Eq. (C.6) will be smaller when superlative PPP output estimates are used for countries. The reason is that the differences in energy-output ratios and carbon-output ratios have been found to be smaller when PPP outputs are used than when MER outputs are used. Hence, the error in the second-term correction factor that will arise from projections based on aggregating individual region or country intensities will be smaller when superlative PPP weights are used since the last term will be smaller.