

Chapter 5

Recommended Methodology for Compiling Poverty PPPs in the 2005 ICP Asia Pacific

Introduction

At the inception stage of the global 2005 ICP, it was generally recognized that support for the ICP would be further enhanced if its results can help improve the methodology for measuring the incidence of poverty in different regions with the use of IPLs. The Global Office of the ICP at the World Bank brought a small group of international experts together to form the PAG, to provide guidance and help set the direction for the work of compiling PPPs for converting IPLs. There was also general agreement that the main thrust of the work during the global 2005 ICP would be to set up the methodology and procedures for the smooth running of the ICP, and that the poverty PPP work would subsequently be added. After carefully assessing the current approach of using consumption PPPs generated by the ICP for converting the IPL and also taking into account what would be feasible within the global 2005 ICP, the PAG recommended a methodology for compiling the PPPs for the conversion of poverty lines. Considering the compilation of poverty PPPs to be an ongoing research that may yield recommendations for future rounds of the ICP, the PAG initiated a number of research projects.¹⁵

This chapter describes the general approach recommended by the PAG and gives details of the implementation of the methodology. The PAG methodology was taken as the baseline methodology

for the poverty PPP study of the 2005 ICP Asia Pacific. The results from implementing the methodology are in Chapter 7. A number of alternative sets of PPPs derived using approaches that deviate from the PAG methodology are also presented in Chapter 7, which also examines the sensitivity of PPPs to the use of different approaches.

PAG's Recommended Methodology

The PAG considered the current practice of simply using PPPs for the consumption aggregate of the ICP as inadequate for the purpose of converting the IPL. The PAG discussed the main issues regarding the ICP consumption PPPs in terms of their commodity coverage and the use of national average weights in deriving the PPPs. While it was generally acknowledged that both the items priced and the weights used would have a bearing on poverty PPPs, the use of incorrect weights to aggregate price data was considered the more immediate problem that had to be addressed. The main focus of the 2005 round is to conduct the ICP effectively, but given the time and financial constraints, it was recognized that it would not be feasible to conduct separate poverty-specific price surveys during this round. Further, the PAG was uncertain about the magnitude of the difference that would be generated if ICP price data were to be substituted for price data from poverty-specific baskets of goods and services.

¹⁵ Additional information on poverty PPP studies is on <http://go.worldbank.org/4YG715RGTO>.

The PAG-recommended methodology for compiling poverty PPPs for the global 2005 ICP had the following elements:

- (i) The price data for poverty PPPs would be the same as that used for the ICP. Therefore, the basic heading PPPs generated from the ICP would be used for computing the poverty PPPs.
- (ii) The weights used in computing the poverty PPPs would be based on the expenditure weights of those households whose expenditure is around the poverty line.
- (iii) The aggregation methodology used would be the same as that used for the ICP, i.e., the Eltetö-Köves-Szulc method for aggregation above the basic heading level.

Although the recommended method appears to be a simple variation of the current ICP methodology, its implementation is complex because it requires the expenditure shares of the poor as weights. The rest of this chapter discusses the main issues and procedures followed in compiling poverty PPPs as part of the 2005 ICP Asia Pacific. The following sections present basic heading PPPs for selected basic headings; describes the compilation of weights for the purpose of aggregation; and provides a brief description of the aggregation methods used at various stages in the poverty PPP study in the 2005 ICP Asia Pacific.

Basic Heading PPPs Based on the 2005 ICP Asia Pacific

The 2005 ICP Asia Pacific price data are the basic input into the computation of poverty PPPs using the PAG methodology. Consistent with the PAG recommendation, basic heading PPPs were drawn from the 2005 ICP Asia Pacific but re-expressed using the Malaysian ringgit as the reference currency. The countries included are Bangladesh, Bhutan, Cambodia, Fiji Islands, India, Indonesia, Lao PDR, Malaysia, Maldives, Mongolia, Nepal, Pakistan, Philippines, Sri Lanka, Thailand, and Viet Nam.

As the poverty PPPs are based on household consumption only, the poverty PPP computations made use of only PPPs for 110 basic headings

that make up the household final consumption expenditure.

Typically, the basic heading PPPs were then combined using the expenditure shares to derive an aggregated PPP. In the case of the poverty PPPs, they were combined using the expenditure share weights for the poor.

Compiling Expenditure Share Weights for the Poor

The main source of expenditure data is the household expenditure surveys conducted regularly in the 16 countries. For the purpose of computing the PPPs, expenditure weights need to correspond to the basic headings used in the 2005 ICP Asia Pacific. As compiling PPPs is not the main objective of the HES, the commodities and groups used in the surveys do not readily correspond with the 2005 ICP Asia Pacific basic headings and categories.

There are two main tasks in the process of compiling weights for the poor. First is to establish correspondence between the 2005 ICP Asia Pacific and HES data sets by constructing a mapping between the two sources. Once such a mapping is established, the second task is to identify the poor households to compute the budget shares and weights for the poor. Identification of the poor is accomplished through an iterative scheme developed by Deaton (2004). Once these two tasks are completed, then a set of weights in the form of expenditure shares for the poor is derived. The two tasks are discussed in more detail in the following paragraphs.

Mapping Household Expenditure Surveys to ICP Basic Headings

Establishing correspondence between HES and the 2005 ICP Asia Pacific data is an arduous, resource-intensive task. Much time and manpower are needed to establish the mappings. Both have been generously allocated by the World Bank in its effort to move this study forward. Basic elements of the approach of Dupriez (2007) are briefly described here.¹⁶

¹⁶ See http://siteresources.worldbank.org/ICPINT/Resources/270056-1195253046582/Dupriez_BuildingaHHCdatabase-fortheCalculationofPovertyPPPs_Mar07.pdf.

Dupriez has written a software based on Stata (versions 8 and 9) to construct HES data sets that can be used for ICP purposes. In the process, he developed a software not only for establishing a mapping of all goods and services in the HES with the ICP basic headings but also for detecting and fixing possible outliers in the data sets. The mapping process consists of the following steps:

- (i) extracting household characteristics,
- (ii) calculating annual consumption for all goods and services covered by the survey,
- (iii) detecting and fixing outliers in consumption values,
- (iv) mapping all goods and services to the corresponding basic headings,
- (v) splitting the values stored in “fake” basic headings, and
- (vi) running quality control tables.

The most important step is the mapping of HES commodities to ICP basic headings. As noted earlier, there are 110 basic headings for HFCE in the 2005 ICP Asia Pacific. Only 107 of those can possibly be obtained from HES. The three basic headings that cannot be mapped are financial intermediation services indirectly measured, purchases by resident households in the rest of the world, and purchases by nonresident households in the economic territory of the country.

Three challenges were encountered in mapping, mainly because HES are not instruments specifically devised for ICP purposes.

- (i) In the HES, several items are like composite commodities that do not exist as basic headings in the ICP. For example, meat could be an expenditure item in the HES in a given country, but corresponding to this are five basic headings within the ICP, e.g., beef and veal; pork; lamb, mutton, and goat; poultry; other meats and meat preparations. In this case, the HES would provide a household expenditure on “meat.” To address this problem, a dummy basic heading (in addition to the 110 basic headings for HFCE) with the title UNBR (unbroken) Meat is constructed. Given that

weights are available for five meat basic headings in the ICP, the total expenditure on UNBR Meat is distributed among the five basic headings on a pro rata basis. (See Dupriez 2007.)

- (ii) Some basic headings in the ICP are aggregates of items in the HES. In this case the solution is simple, and the relevant aggregates are formed out of the individual components.
- (iii) No data corresponding to certain basic headings in the ICP are available in the HES. For example, in Bangladesh, 29 basic headings in the ICP had no corresponding entries in the HES. These basic headings accounted for only 1.2% of the total ICP weights. The corresponding figures for Indonesia, Sri Lanka, and Thailand, respectively, are 15 (0.53%), 11 (2.6%), and 6 (4.26%).

Identifying Households on the Poverty Line

The next task in compiling expenditure share data is to identify all households in different countries that are deemed to be on the poverty line. There are several issues to resolve here and two steps to follow.

Step 1. Select a poverty line.

The first step is to select a poverty line for the purpose of extracting the expenditure patterns of the poor. Several choices are available, and the choice made is likely to have an effect on the numerical values of the PPPs generated. In this study two poverty lines are selected for the purpose of examining the sensitivity of poverty PPPs. One poverty line is a variant of the \$1/day IPL, and the other is the national poverty line used in Indonesia. The Indonesian poverty line for 2005 was set at Rp1,549,296/annum.¹⁷ The reasons for the selection of the Indonesian poverty line are discussed in Chapter 7. For purposes of exposition, the following discussion is based on the IPL of \$1/day.

The \$1/day IPL in recent literature refers to \$1.08 in 1993.¹⁸ For the purpose of this study, the IPL for 2005 is required. The \$1.08 IPL in 1993

¹⁷ Data provided by Badan Pusat Statistik.

¹⁸ See Chen and Ravallion (2004) for a description of the historical development of this approach.

was converted to 2005 prices using the CPI in the United States. The new line for 2005 is found to be \$1.46. This means that over the period 1993 to 2005, consumer prices in the US increased by 35%.

To identify the households around a given poverty line, the IPL of \$1.46 needs to be converted into local currency units. Suppose Malaysia is selected as the country of interest. As a starting point, PPPs for the HFCE for Malaysia from the recently released Final Results of the Global ICP¹⁹ indicate a PPP of \$1 = RM2.11. Then the IPL in RM is equal to $1.46 \times 2.11 = \text{RM}3.0806/\text{day}$. (The \$ poverty line can be similarly converted into any other national currency unit using an appropriate PPP.) This can be annualized by multiplying 3.0806 by 365 giving a PPP-based annual poverty line of RM1,124.42. Therefore, any person with expenditure less than the poverty line of RM1,124.42 is considered poor.

Step 2. Identify households around the poverty line.

Once a poverty line is selected, the next step is to identify the households on the poverty line. As the poverty line for Malaysia used in the illustration is RM1,124.42, a single figure in monetary terms, it is likely that no individual in the Malaysian HES would have an annual expenditure exactly equal to RM1,124.42. It is necessary to use a small interval around the poverty line for the purpose. A small bandwidth²⁰ is selected so as to define this interval. In the current study, the following formula is used for computing the bandwidth. The bandwidth, h , is given as

$$h = 1.059sn^{-1/5} \quad (1)$$

¹⁹ Final Results from the Global ICP are available from the World Bank (see http://siteresources.worldbank.org/ICPINT/Resources/ICP_final-results.pdf).

²⁰ This is a term used in Kernel smoothing. Deaton (2004 and 2006) describes the use of the Kernel smoothing technique as a solution to this problem.

Table 7. Bandwidths and Estimated Population Size in Intervals around the Poverty Line

Country	National Poverty Lines	h	Poverty Line \pm h	
			Bandwidth	
			Lower Bound	Upper Bound
Bangladesh	9,672.72	203.73	9,468.99	9,876.45
Bhutan	8,884.32	844.92	8,039.40	9,729.24
Cambodia	775,260.00	29,969.10	745,290.90	805,229.10
Fiji Islands	1,820.00	89.95	1,730.05	1,909.95
India	4,905.63	119.12	4,786.51	5,024.75
Indonesia	1,549,296.00	34,908.92	1,514,387.08	1,584,204.92
Lao People's Democratic Republic	1,115,520.00	127,469.94	988,050.06	1,242,989.94
Malaysia	1,860.00	145.82	1,714.18	2,005.82
Maldives	5,475.00	1,131.39	4,343.61	6,606.39
Mongolia	296,808.00	16,949.40	279,858.60	313,757.40
Nepal	7,695.60	587.02	7,108.58	8,282.62
Pakistan	10,543.68	183.12	10,360.56	10,726.80
Philippines	14,046.00	535.19	13,510.81	14,581.19
Sri Lanka	21,804.00	830.19	20,973.81	22,634.19
Thailand	14,904.00	901.82	14,002.18	15,805.82
Viet Nam	2,076,000.00	122,265.93	1,953,734.07	2,198,265.93

h = bandwidth.

Note: National poverty line refers to the latest available poverty line in local currency units.

where s is the sample standard deviation and n is the sample size.²¹

In practice, expenditure weights are computed using household expenditure data where total household expenditure is recorded in the survey. Therefore, in applying the formula, n is the total sample size used in the HES.

In the case of Malaysia, the latest household expenditure data available refer to the year 2004 and the corresponding value for the bandwidth is calculated to be $h = \text{RM}145.82$ and therefore, the interval around the poverty line of $\text{RM}1,124.42$ is given as 978.6 and 1,270.24. In calculations in the sensitivity analysis, a smaller bandwidth of 0.5 h is used. A smaller bandwidth means that a smaller

number of households will belong to the given interval whereas a wider bandwidth will capture households further away from the poverty line.

Once the bandwidth and the interval around the poverty line are determined, it is then possible to identify all the households whose per capita expenditures lie within the interval considered. Table 7 presents bandwidths computed using equation (1).

Computing the Expenditure Share Weights

For a given household r , the expenditure share of commodity²² n is given by

$$w_{nr} = \frac{P_{nr}q_{nr}}{\sum_{n=1}^N P_{nr}q_{nr}} \quad (2)$$

²¹ The bandwidth, h , given by the formula is optimum in the sense that it minimizes the mean square error when the Kernel used has a standard normal probability density function. For more details, refer to any standard text on nonparametric methods. Greene (2003, 453–56) provides a simple exposition of the methodology used here. This is also similar to the bandwidth concept used in Deaton (2006).

²² The term “commodity” is used in place of basic headings, which are more relevant in PPP calculations.

Table 7. Bandwidths and Estimated Population Size in Intervals around the Poverty Line (continued)

Poverty Line $\pm h$		Poverty Line $\pm 0.5 h$			
Sample Size		Bandwidth		Sample Size	
Households	Headcount	Lower Bound	Upper Bound	Households	Headcount
342	1,701	9,570.85	9,774.58	167	844
279	1,659	8,461.86	9,306.78	132	769
806	3,895	760,275.45	790,244.55	405	1,951
230	1,034	1,775.02	1,864.98	127	572
2,248	14,013	4,846.07	4,965.19	1,165	7,134
1,709	7,798	1,531,841.54	1,566,750.46	831	3,819
1,346	8,803	1,051,785.03	1,179,254.97	668	4,323
604	3,551	1,787.09	1,932.91	316	1,818
126	903	4,909.31	6,040.69	67	484
741	3,504	288,333.30	305,282.70	366	1,767
335	1,836	7,402.09	7,989.11	170	950
463	3,235	10,452.12	10,635.24	230	1,579
1,708	8,229	13,778.41	14,313.59	850	4,120
924	3,991	21,388.90	22,219.10	464	1,978
1,719	7,014	14,453.09	15,354.91	840	3,411
464	2,243	2,014,867.03	2,137,132.97	233	1,134

where p_{nr} represents the price of n -th commodity ($n=1,2,\dots,N$) for household r and q_{nr} represents the quantity of n -th commodity consumed by household r . The expenditure share shows the proportion of total expenditure of household r for commodity n . Therefore, it is easy to compute household expenditure share weights for each of the households.

Suppose expenditure share weights are computed for each household r in a total of R households. To compile weights to be used in the PPP computation, the weights need to be aggregated to derive the necessary weighting scheme. Two types of weighting schemes are considered, *democratic* and *plutocratic weights*.

Democratic weights are simple averages of the expenditure shares of all households belonging to the income interval defined around the poverty line. If R households belong to this group, then the democratic weight based on these R households is given as:

$$w_n = \frac{1}{R} \sum_{r=1}^R w_{nr} \quad (3)$$

where w_n is the expenditure share of n -th commodity that will be used as a weight in the PPP computations.

These weights are considered democratic as each household gets the same weight in the computation of the expenditure share weights.

Plutocratic weights are weights that represent the whole set of households as a group. These weights are derived by considering the total expenditure of all the households on a given commodity, n , as a share of the total expenditure on all commodities by all households. Therefore, the whole group of households is treated as a single entity. Plutocratic weights are calculated as

$$w_n = \frac{\sum_{r=1}^R p_{nr} q_{nr}}{\sum_{r=1}^R \sum_{n=1}^N p_{nr} q_{nr}} \quad (4)$$

The numerator of this expression shows the total amount spent on commodity n by all the R households taken together. In contrast, the

denominator shows the total expenditure of all the households on all the commodities.

It is easy to see that in the computation of plutocratic weights, households with larger expenditures tend to be given a higher weight in the computation of the expenditure shares.

While democratic weights are the preferred choice, plutocratic weights are also used in PPP computations to check the sensitivity of the derived PPPs to the choice of the method of deriving average expenditure share weights.

The following numerical example illustrates the concepts of democratic and plutocratic weights and shows the kind of average weights that can result from their use.

	Household 1		Household 2	
	Expenditure	Budget share	Expenditure	Budget share
Item 1	50	0.33	20	0.67
Item 2	100	0.67	10	0.33
	Democratic weights		Plutocratic weights	
Item 1	0.5		0.38	
Item 2	0.5		0.62	

This example illustrates the two different methods of computing averages across households. Democratic weights offset the unbalanced patterns of the two households whereas plutocratic weights tend to retain the expenditure pattern of the richer household whose total expenditure is 150 (household 1) compared with only 30 by household 2.

In practice, however, the use of democratic and plutocratic weights is likely to produce less extreme results than the example above, as the averages are taken over all the households around the poverty line, which are expected to have similar total expenditures. The sensitivity of PPPs to the use of democratic and plutocratic weights is further examined in Chapter 7.

Addressing Nonavailability of HES for the Benchmark Year

Because HES are costly to conduct, countries conduct them once in 3–5 years. Further, the HES are not perfectly synchronized across the countries participating in the poverty PPP study. Considerations

concerning the use of HES data that may not correspond to the benchmark year, in this case 2005 for the ICP Asia Pacific, should be identified.

Table 8 shows the HES data used in all the PPP computations in the study. The idea was to use the HES data closest to the 2005 benchmark year. However, there were two exceptions. At the time of the analysis, only 2001 HES data for Pakistan was available. Information from 2001 was processed before data for a more recent year became available. The situation with Indonesia is also the same. After the initial processing of the 2002 HES data, 2005 data became available. As the processes involved in the analysis are complex, it was not possible to make use of the 2005 HES data. Details of the HES from different countries used in this study are presented in Appendix Table 2.

In Table 8, column (2) gives the equivalent of \$1/day (\$1.08 in 1993, which is equivalent to \$1.46 in 2005) converted into the respective local currency units using PPPs from the 2005 ICP Asia Pacific and the Global ICP. The column shows what may be considered the conversion of the IPL to local currency units.

To use that poverty line to compute expenditure weights discussed in the previous section, it is necessary to find all households around this poverty line in each of the countries with the use of the latest available HES data. In the case of Bangladesh, India, and Mongolia, HES data are available for 2005 and therefore, the figures in columns (2) and (3) can be used directly for these countries. In the case of Malaysia, however, the latest available HES data refer to 2004. To find all those households around the poverty line, it is necessary to adjust the poverty line of RM3.08 in 2005 to 2004 prices. The resulting poverty line is RM2.99. This adjustment is made using the CPI data drawn from *International Financial Statistics 2007* of the International Monetary Fund. Similarly, the poverty line for Viet Nam is dong (D) 7,843.08, which is adjusted downward from the 2005 level of D8,491.70. In the case of Pakistan the available HES is for 2001 and therefore, its poverty line of Pakistani rupees (Prs)30.14 is adjusted to Prs24.20, reflecting the price changes over the period 2001 to 2005.

The assumption here is that the expenditure patterns, in particular those of the poor, do not change too dramatically over a short period and that

Table 8. Household Expenditure Survey Availability and Consumer Price Index Movements

Country	\$1/day equivalent		HES			
	2005	2005	2004	2003	2002	2001
	(1)	(2)	(3)	(4)	(5)	(6)
Bangladesh	36.60	36.60				
Bhutan	26.11			23.71		
Cambodia	2,403.90			2,189.10		
Fiji Islands	2.00				1.83	
India	23.21	23.21				
Indonesia	6,132.80				4,903.30	
Lao People's Democratic Republic	5,851.32			4,942.83		
Malaysia	3.08		2.99			
Maldives	12.79				12.92	
Mongolia	739.16	739.16				
Nepal	37.18			33.84		
Pakistan	30.14					24.20
Philippines	33.95			29.76		
Sri Lanka	55.07				43.13	
Thailand	23.95				21.90	
Viet Nam	8,491.70		7,843.08			

HES = household expenditure survey.

expenditure share patterns extracted from the HES year would be equally applicable to the benchmark year.

Computing the PPPs

An iterative scheme is used in computing the PPPs. Once the expenditure share weights are compiled, then the basic heading PPPs can be combined with these weights to derive a new set of PPPs. Once new PPPs are obtained, they can, in turn, be used in deriving a new set of weights and the process is continued until the PPPs converge completely. However, combining price data and expenditure share weights requires the use of an appropriate aggregation methodology.

Aggregation Methods

As there are 16 countries participating, it is necessary to make use of multilateral index numbers in deriving PPPs. Multilateral comparisons essentially involve comparisons of prices or quantities between every pair of countries included in the study. If there are 16 countries, then there will be 120 binary comparisons involving pairs of countries, and every binary comparison is considered equally essential. This is particularly the case with international comparisons undertaken by international organizations. In multilateral comparisons, it is necessary to ensure internal consistency in the results reported; therefore, the methods used for this purpose are expected to satisfy some basic properties.

Desirable Properties of PPPs

Suppose PPPs are computed for the currencies of all the countries involved using a particular index number formula. To meaningfully express PPPs, it is necessary to nominate a country whose currency, j , is taken as the reference currency.²³ By definition, the PPP of the reference currency is always equal to 1. Hence, PPP_j represents the PPP of currency j , expressed in the units of a reference currency.

Base-country invariance is an important property of PPPs. It means that the results of comparisons do not depend on the choice of the base country or of the numeraire currency. This property ensures that all the countries involved in

²³ A basket of currencies may also be nominated to serve as a reference currency for the purpose of PPP computations.

the comparisons are treated symmetrically, which reflects a level of neutrality expected of international comparisons.

Transitivity is a crucial property of PPPs because it ensures that comparisons made between any pair of countries are mutually consistent. Having PPPs that are transitive means that an indirect comparison between two countries via a third country will yield the same result as a direct comparison between the first two countries. For example, if the countries concerned are j , k , and l , the PPP between countries j and k ($PPP_{j,k}$) will be identical to the product of the PPP between j and l ($PPP_{j,l}$) and the PPP between l and k ($PPP_{l,k}$). This relationship can be expressed algebraically as follows:

$$PPP_{j,k} = PPP_{j,l} \times PPP_{l,k} \quad (5)$$

In more general terms, transitivity is satisfied if PPPs are such that the above equation holds for any selected set of three countries, say j , k and l (1, 2, ..., M).

Characteristicity is an important property in the case of multilateral comparisons that satisfy the transitivity property. Transitivity implies that a comparison between countries j and k (for example, between Malaysia and India) would be affected by country l (say Thailand). Characteristicity stipulates that a formula should minimize possible distortions to binary comparisons created due to the transitivity restriction.

Additivity and other properties are also desirable. For example, the index number formula should be such that PPPs derived are invariant to the choice of the units in which quantities are measured. Whether the price of rice is measured per one kilogram unit or per 10 kilogram units, the resulting PPP should be the same. This result is usually referred to as “axioms.” Details of the axiomatic approach can be found in Diewert (1988) and Balk (1995). Of all such properties, additivity is one of the most commonly discussed properties. Additivity ensures that subaggregates converted using PPPs add up to a total GDP converted into a real aggregate using a PPP at the GDP level. Additivity is a property that is obviously satisfied by the national accounts in local currency units. If PPPs are derived using an aggregation method that guarantees additivity, such as the Geary-Khamis (GK) method, the national accounts expressed in a common currency unit will also satisfy additivity. The estimates in this

publication are compiled using mainly the EKS method, which is not an additive method.

Two scenarios are likely to be considered when aggregation methods are being selected for PPP compilation. If the basic heading PPPs are available from the 2005 ICP Asia Pacific, then the only step involved in poverty PPP compilation is to aggregate such basic heading PPPs with the use of expenditure share weights derived using the iterative procedure described earlier. If the price data are obtained from a price survey directly, then it is necessary to aggregate item-level price data to derive basic heading PPPs. At this level, no weights are available. In the interest of completeness, methods of aggregation below and above the basic heading level are presented here.

Aggregation below the Basic Heading Level

For this round of international comparisons, the main aggregation procedure recommended for deriving PPPs at the basic heading level from item-level price data is the country-product-dummy (CPD) method.²⁴ The basic input into the CPD method is the national average price for each of the products.

In compiling the poverty PPP, the CPD method was used in aggregating price data collected from the poverty-specific price surveys conducted in the 16 participating countries. The starting point for the CPD approach, therefore, is a matrix of prices (in local currency) for priced products within each of the 16 countries. As expected, there were gaps in the price matrix because it was not possible (nor generally desirable) for all countries to price every product on the list.

The CPD method is a regression technique. The underlying model is multiplicative and assumes that prices vary by product within countries at the same rate across all countries, and that prices vary between countries at the same rate across all products. This is sometimes referred to as the *law of one price*. As is usual with a regression equation, an error term (also multiplicative in this case) is required to handle variations in the observed product/country prices from those generated by the model. In practice, one country and one product in one country have to be

chosen as the base and all other product-country combinations are measured in terms of their variation from these bases.

The multiplicative CPD model can be shown using a simple example. Assume that we have m countries and their product list consists of n products. Then, for each product (i) in each country (j), the price observed is p_{ij} for $i = 1, 2, \dots, N$ and $j = 1, 2, \dots, M$. This means that there are M countries and N commodities. In this case $M=16$ and $N=155$. Note that the prices p_{ij} are expressed in each country's national currency. The CPD model is expressed as $p_{ij} = \alpha_i \beta_j v_{ij}$ where α_i is the product term, β_j is the country term, and v_{ij} is the error term.

Additive models have some useful properties and so, in practice, the CPD model is converted from a multiplicative one to an additive one by expressing the terms in the model as logarithms:

$$\begin{aligned} \log(p_{ij}) &= \log(\alpha_i \beta_j v_{ij}) \\ &= \log(\alpha_i) + \log(\beta_j) + \log(v_{ij}) \quad (6) \\ &= \eta_i + \pi_j + u_{ij} \end{aligned}$$

This method is known as CPD because the model can be rewritten using dummy variables for commodities and countries. The model is

$$\begin{aligned} \ln p_{ij} &= \pi_1 D_1 + \pi_2 D_2 + \dots + \pi_M D_M + \eta_1 D_1^* \\ &+ \eta_2 D_2^* + \dots + \eta_n D_n^* + u_{ij} \quad (7) \end{aligned}$$

where D_i is the dummy variable for commodity i taking the value 1 if the price observation refers to commodity i , and the value 0 otherwise; D_j^* is a country dummy variable that takes a value of 1 if the price observation refers to country j , and a value of 0 otherwise.

The parameters in the model are estimated using a least-squares approach. Given that the model requires the outputs to be expressed in terms of one country's currency, the outputs are simply PPPs expressed in terms of that base country. In the model, if we assume that the base country is country 1, then $\alpha_1 = 1$. In addition, it is necessary to select a product to act as a base product, so if we make $\beta_1 = 1$, then the model produces estimates of prices in terms of their variation from product 1 in country 1. Any other country can be made the base country simply by dividing each country's PPP by the new country's

²⁴ A popular alternative to the CPD method is the EKS method used by Eurostat and OECD. A comparative assessment of the relative merits of these methods is available in the *ICP Handbook* (World Bank 2007). Rao (2004) discusses all aspects of the CPD method including a comparison with the EKS method.

PPP. The CPD model assigns the same weight to each product's price, so it is often referred to as an "unweighted model," although it should really be described as a model with equal weights.²⁵

One useful output from the CPD model is a set of estimated prices for each product for each country. These prices provide an estimate of what the prices would be if the relationships set out above held in practice. The differences between observed prices and these modeled prices can provide an indication of possible problems with the prices provided by a country. Large differences indicate possible problems, e.g., because the prices for the same product vary significantly across countries or because the relativities between prices of products within a country vary significantly compared with those in other countries. The distribution of these differences provides the underlying basis for the Dikhanov table, developed in the World Bank as an editing tool (see Part 4 of ADB 2007b). The distributions can be graphed to provide a simple means of identifying potential problem prices, either for a particular product or within a country.

The CPD method is used in deriving PPPs at the basic heading level in the course of processing price data collected through poverty-specific price surveys.

A variant of the CPD method, known as the country-product-representativity-dummy (CPRD) method, can use any additional information that may be available indicating the representativity status. The method introduces a dummy variable in the CPD regression, R_{ij} , for commodity i in country j such that if the product is representative, then $R_{ij} = 0$; and if it is not representative, then $R_{ij} = 1$. Then the CPD model presented before is extended to include the representativeness dummy variable, R_{ij} . The new model is then given as

$$\log(p_{ij}) = \log(\alpha_j) + \log(\beta_j) + \delta \log(R_{ij}) + \log(v_{ij}) \quad (8)$$

where the new parameter δ captures the effect of inclusion of price data for nonrepresentative items. Thus, the CPRD model is supposed to provide estimates of PPPs, which are more reliable

than the PPPs that do not take into account unrepresentativeness of some of the price data used.

In the poverty PPP study, the CPRD method could not be used as the data on representativity provided by the participating economies had not been very reliable. Use of incorrect indications of representativity or nonrepresentativity is likely to introduce further biases into the results.

Further details of the CPD and CPRD methods can be found in the *ICP Handbook* (World Bank 2007).

Aggregation above the Basic Heading Level

It is easier to describe the foregoing methods using the following notation. Let p_{ij} and q_{ij} denote, respectively, the price and quantity of product i in country j . Since the aggregation here makes use of basic heading PPPs, p_{ij} simply denotes the basic heading PPP for i -th basic heading in j -th country. The quantity q_{ij} refers to the basic heading that is an aggregate over several commodities. Therefore, q_{ij} is given by e_{ij}/p_{ij} where the price refers to the basic heading PPP of commodity i in country j .

$$\text{Paasche: } P_{jk} = \frac{\sum_{i=1}^N p_{ij} q_{ij}}{\sum_{i=1}^N p_{ik} q_{ij}} \quad (9)$$

$$\text{Laspeyres: } L_{jk} = \frac{\sum_{i=1}^N p_{ij} q_{ik}}{\sum_{i=1}^N p_{ik} q_{ik}} \quad (10)$$

$$\text{Fisher: } F_{jk} = \sqrt{P_{jk} L_{jk}} \quad (11)$$

$$\text{Tornqvist: } T_{jk} = \prod_{i=1}^N \left[\frac{p_{ik}}{p_{ij}} \right]^{\frac{(w_{ij} + w_{ik})}{2}} \quad (12)$$

where: $w_{ij} = \frac{p_{ij} q_{ij}}{\sum_{k=1}^N p_{kj} q_{kj}}$ are expenditure shares.

For the purpose of poverty PPP compilation, the total expenditure is taken to be equal to the

²⁵ There are weighted versions of the CPD model. For some important applications of weighted CPD models, see Rao (2005) and Diewert (2005).

poverty line expressed in the respective local currency units. The expenditure share weights are the “average” weights derived using data for all the households within a given bandwidth around the poverty line.

For the empirical results on PPPs reported in Chapter 7, three different aggregation procedures were used. These are the EKS method, the weighted CPD method, and the GK method. The main reason for using these three different methods is to examine the sensitivity of the results to the use of different index number methods. In this poverty PPP study, the CPD method is used for aggregating prices below the basic heading level. The EKS method is used for aggregation above the basic heading level. However, the sensitivity analyses reported in Chapter 7 also consider the GK and the weighted CPD methods. The general recommendation for the ICP is to use the EKS method.

The EKS Method. The computational form for the EKS index is given by

$$PPP_{jk} = \left[\prod_{l=1}^M [F_{jl} \cdot F_{lk}] \right]^{1/M} \quad (13)$$

where F_{jk} denotes the Fisher price index number for country k with country j as the base and M denotes the total number of countries. The Fisher index formula is given earlier in equation (11).

The EKS method recognizes that the Fisher binary index numbers, which are the commonly preferred binary index numbers, do not satisfy the property of transitivity, and, therefore, are not suitable for use in multilateral comparisons. The EKS method generates a multilateral index number using the binary Fisher index numbers as building blocks. Two useful properties of the EKS are to be noted. First, the EKS method maintains characteristicity by ensuring that the EKS index numbers deviate the least from the binary Fisher index numbers. Second, the EKS has a simple and intuitive interpretation that a binary comparison between countries j and k is an unweighted average of all the linked comparisons between j and k using links $l = 1, 2, \dots, M$.²⁶

The GK Method. This method has been used in most international comparisons until recently. Kravis et al. (1982) give an excellent exposition of the GK method. The method provides a framework for

computing PPPs and international average prices, P_s , of commodities. The GK method defines these using the following interrelated equations:

$$PPP_j = \frac{\sum_{i=1}^n p_{ij} q_{ij}}{\sum_{i=1}^n P_i q_{ij}} \quad (14)$$

and

$$P_i = \frac{\sum_{j=1}^M (p_{ij}/PPP_j) q_{ij}}{\sum_{j=1}^M q_{ij}} \quad (15)$$

where P_i 's are international prices.

It is clear from the equation for P_i that the international price of a commodity is defined as a weighted average of national prices, with weights proportional to the quantities. The PPPs from the GK method are computationally derived using an iterative scheme that solves the system of interrelated equations for the international average price, P_p , and the purchasing power parities, PPP_j .

The preference for the GK method is mainly due to its intuitive simplicity and the property of additivity it satisfies. It ensures that the real aggregates for different components of the GDP sum up to the real GDP. However, the definition of international prices has a tendency to be influenced by prices observed in richer and larger countries. This tendency induces bias, which tends to overstate the real incomes of the poorer economies, an effect similar to the Gerschenkron effect.²⁷ In recent years, there has been a movement away from the GK method and it is being increasingly replaced by the EKS and the weighted CPD methods.

Weighted CPD Method. The standard CPD formulation is a simple regression model that regresses the logarithm of observed prices on a set of dummy variables representing the commodity and country to which a given price observation refers. It uses the following regression model:

$$\ln p_{ij} = \pi_1 D_1 + \pi_2 D_2 + \dots + \pi_M D_M + \eta_1 D_1^*$$

²⁶ See Rao (2001) for further details on the EKS method.

²⁷ See the recent work of Dowrick and Akmal (2005) for more details of the GK method.

$$+ \eta_2 D_2^* + \dots + \eta_n D_n^* + u_{ij} \quad (16)$$

where D and D^* , respectively, refer to country and product dummy variables. This formulation was used in handling missing observations. It was also used as a method for aggregating price data below the basic heading level.

Rao (1995) showed that it is feasible to consider a generalization of the CPD method that can be used for aggregation above the basic heading level. Rao proposed an extension that allows for the use of weights—an extension with its roots in weighted least squares—with weights reflected by the expenditure shares. The model is equivalent to running the following regression model with transformed observations.

$$\begin{aligned} \sqrt{v_{ij}} \ln p_{ij} = & \pi_1 \sqrt{v_{ij}} D_1 + \pi_2 \sqrt{v_{ij}} D_2 + \dots \\ & + \pi_M \sqrt{v_{ij}} D_M + \eta_1 \sqrt{v_{ij}} D_1^* + \dots \\ & + \eta_n \sqrt{v_{ij}} D_n^* + u_{ij} \end{aligned} \quad (17)$$

The required PPPs are simply given by $PPP_j = \exp(\hat{\pi})$ where $\hat{\pi}_j$ is the least-squares estimator of $\hat{\pi}_j$.

Even though the weighted CPD is a simple extension of the CPD model, Rao (2004 and 2005) demonstrated that it is a powerful technique with very important properties, a few of which are listed below.

- (i) The weighted CPD method is equivalent to the Rao (1990) system for multilateral comparisons. This result, proven in Rao (2005), establishes a link between the econometric approach and the standard PPP and international price approach in the GK method. Since the weights here are based on shares, this method is size-neutral.
- (ii) When applied to binary comparisons, the method results in index numbers that are superlative. In fact, it is possible to obtain an explicit form in the case of binary comparisons, and the resulting index is a geometric average of price relatives (similar to the Tornqvist index). Diewert (2005) shows that a number of other indices can also be generated by varying the CPD model specification.

(iii) The weighted CPD model allows for a more complex specification of the disturbances. For example, spatial structures in price relatives can be gainfully exploited and incorporated into the computations through a spatially autocorrelated disturbance specification (see Rao 2004 for more details).

(iv) The CPD model can be viewed as a simplified hedonic model. This interpretation suggests the possible use of the generalized CPD method to include quality characteristics explicitly in the regression specification. Incorporation of quality and outlet characteristics allows for the use of a more general approach to international comparisons where the standard approach of pricing very tightly specified items can be replaced by an approach based on loosely specified items with all the specifications recorded for each price observation. As a result of this approach, it is possible to improve the representativeness of the baskets priced for purposes of the ICP.

All the three aggregation procedures—EKS, GK, and weighted CPD methods—have been used in Deaton (2006). A comparative analysis of the results from the three methods provides a measure of the sensitivity of PPPs to the aggregation method used.

The multilateral index number methods are designed to take into account differences in expenditure patterns of countries. For example, the EKS method is built from binary Fisher PPPs such that the multilateral PPPs from the EKS method deviate the least from the binary PPPs. Similarly, the weighted CPD method explicitly accounts for differences in expenditure patterns. However, it may be noted that the reliability of PPPs increases when comparisons involve similar countries with similar expenditure patterns.²⁸ In cases where countries are very dissimilar, it may be necessary to use a sub-regionalized approach.

²⁸ There are multilateral methods (see Rao and Timmer 2003) that improve upon the EKS method by explicitly accounting for dissimilarities in expenditure patterns in the computation of multilateral PPPs. But these methods are still not widely used.

Conclusion

The main objective of this chapter has been to articulate the methodology recommended by the PAG at the Global Office of the ICP and then discuss the implementation strategy for the recommendations. It is clear from this chapter that the PAG recommendation is only a simple first step in the quest for improved PPPs for estimating poverty and that the actual implementation is a lot more complex

than what is involved in the general ICP approach in computing PPPs. The main complexity is introduced because the reference population representing the poor is not determined independently of the poverty PPPs that rely on the expenditure patterns of the poor. Therefore, it is necessary to simultaneously determine the poor and the poverty PPPs. This is achieved through the iterative scheme proposed by Deaton (2004 and 2006) and implemented here using household expenditure data for the 16 countries.