

Validation of Household Consumption Data

The national coordinating agencies (NCAs) of the economies participating in the regional comparisons of the International Comparison Program (ICP) provide the regional coordinating agencies (RCAs) with a set of national quarterly and annual purchasers' prices for items chosen from a common list of precisely defined items. These regional lists include both regional and global items belonging to the global core list (GCL). The quarterly prices refer to the quarterly national average prices; the annual purchasers' prices refer to the annual national average prices of the year of the comparison. Both sets of prices cover the whole range of final goods and services included in the gross domestic product (GDP). These prices are used to calculate purchasing power parities (PPPs) for basic headings (BHs). The basic heading PPPs are further used to derive intraregional measures of price and volume relatives for the economies participating in the comparison. The regional PPPs are then linked to form a global set of PPPs and measures of price and volume relatives. To complete the process, the measures are published by the RCAs and the World Bank, thereby reaching a variety of users, including policy makers, economic analysts, researchers, politicians, and journalists, as well as the general public. It is therefore essential that the prices on which the PPPs are based are rigorously checked and corrected for error—that is, validated—before the final PPPs are calculated.

Only after this step can the measures contribute accurately to informed debate. Validation of price data is thus an important phase in producing PPPs.

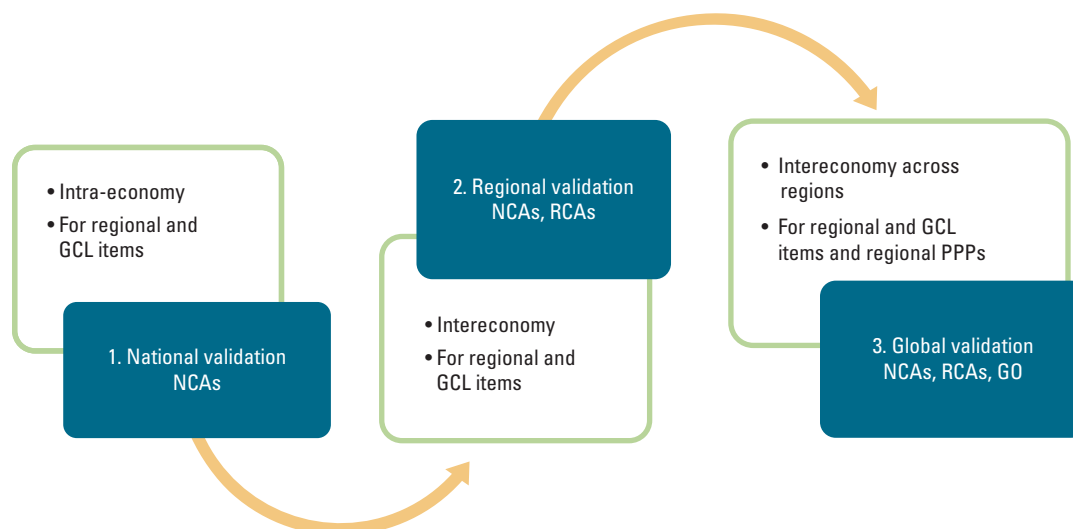
As a multilateral comparison, the ICP exercise requires not only correct but also comparable price data. Correctness and comparability require in turn that the validation process be supported by close cooperation and collaboration among the national and regional agencies and the Global Office. This chapter provides a step-by-step breakdown of the complete validation process for the household consumption survey.

The ICP validation process consists of three stages: (1) intra-economy validation at the national level; (2) intereconomy validation at the regional level; and (3) global validation at the global level. Figure 15.1 illustrates the stages, and the annex to this chapter summarizes the phases and steps in each stage.

Intra-economy validation is directed at an economy's individual price observations and the resulting average prices. The objective is to verify that price collectors within an economy have priced comparable products and have priced them correctly. It is carried out mainly by the economy's NCA. There may or may not be input from the RCA, depending on the region.

Intereconomy validation takes place after intra-economy validation and is carried out by

Figure 15.1 Validation Stages: National, Regional, and Global Levels, ICP 2011



Source: ICP, <http://icp.worldbank.org/>.

Note: NCA = national coordinating agency; RCA = regional coordinating agency; GCL = global core list; GO = Global Office.

the NCAs and RCAs, but with overall coordination by the RCAs. Intereconomy validation is directed at the average prices reported by the participating economies, or in some cases at the individual price observations and the price ratios that the prices generate between the economies. The objective is to verify that price collectors in different economies have correctly priced comparable products.

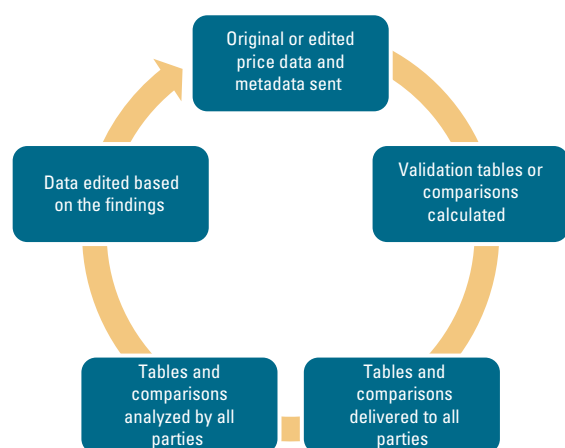
The final validation stage is the global validation; the Global Office and the NCAs and RCAs validate prices for the regional and GCL items, regional PPPs, and subsequent global PPPs. The objective is to ensure that prices collected for the items across regions are comparable and that the linked global PPPs are plausible in wider terms.

Prices for consumer products can be collected on a monthly or quarterly basis, with the final goal of producing annual average prices. Validation at the economy level should be carried out following the price collection schedule on either a monthly or a quarterly basis. Validation at the regional level should be carried out according to the agreed validation schedule, preferably on a quarterly basis. The global-level validation should be carried out on a quarterly basis. Various phases of validation require careful planning of the overall

validation timetable. In addition, timeliness is a crucial aspect requiring serious consideration, first because of the need to release data on a timely basis and second because the longer the delay between price collection and verification, the more difficult it becomes to correct erroneous prices. If economies within a region are not able to follow a common timeline for the validation, the RCAs may need to create validation groups for the regions' economies.

The intereconomy validation is an especially iterative process requiring a number of rounds of editing and verification. Possible errors are uncovered by identifying prices that diverge significantly from other prices in the series. They are detected by having a measure of divergence that is greater than a given critical value or a value that falls outside a given range of acceptable values. Divergence measures are generally defined by the parameters of the series being edited; examples are the average and the standard deviation. Thus if some of the possible errors identified in the initial edit are found to be actual errors and are corrected, the parameters of the price series will change and so will the divergence measures of each price remaining in the series. A second edit will find new possible errors that will need to be verified. Again, when actual errors are

Figure 15.2 Validation Iterations, ICP 2011



Source: ICP, <http://icp.worldbank.org/>.

corrected, the parameters of the price series will change, which may lead to more possible errors being detected if a third edit is done. Usually the number of new possible errors falls as validation progresses, until the return on further rounds is considered marginal and not worth pursuing. Figure 15.2 illustrates validation iterations.

The following sections deal with each validation stage individually. Stages are divided into validation phases, which are further divided into the validation steps to be carried out. But first, what are the objectives and related general means of validation, as well as the limits faced?

VALIDATION OBJECTIVES, MEANS, AND LIMITS

The *first objective* of the validation process is to rid data of pricing errors. Two types of nonsampling errors can be identified: price error and product error.

- A *price error* occurs when price collectors price products that match the item specification, but they fail to record the price correctly. A price error can also occur during the process of reporting and transmitting a price that initially was recorded correctly. Associated with each price is a quantity: the specified quantity (the quantity to be priced) and the reference quantity (the quantity to which the price collected is to be adjusted). A price error can also arise because, even though the price

is recorded correctly, the quantity priced is recorded incorrectly (or is recorded correctly and the error is introduced later, during processing). Thus the adjusted price for the reference quantity, which is the price validated, will be wrong as well.

- A *product error* occurs when price collectors price products that do not match the item specification and neglect to report having done so. Perhaps they are not aware of the mismatch, such as when the item specification is too loose, or perhaps they price a substitute product as required by the pricing guidelines but do not mention it on the price reporting form. Price collectors are usually instructed to collect the price of a substitute product if they are unable to find the specified product. They are further instructed to flag the substitution and to note the differences between the substitute product and the specified product. Flagging brings the substitution to the attention of the NCA, which, together with the RCA, can then decide what to do with the price collected. It may be possible to adjust the price for quality differences between the product priced and the product specified. Alternatively, if other economies report prices for the same substitute product, price comparisons can be made for the substitute product as well as for the product originally specified. If neither one of these options is possible, the price will have to be discarded. Substitution does not in itself introduce an error; it is the failure of price collectors to flag and document the substitution that leads to a product error.

Price and product errors are detected by identifying problematic observations or outliers. These are observations that diverge so much from the average that they are treated, at first glance, as implausible and therefore in need of further investigation and verification. Outliers are not rejected automatically; rather, they are investigated to determine whether they are genuine problematic observations. Indeed there is no exact rule about what makes a price an outlier, and so an element of judgment is required. The amount of divergence that triggers further verification depends largely on the variance of the price observations. It is also

possible that the high price variation reflects the economy's situation correctly. In these circumstances, the RCA must rely on the economy's experience and local market knowledge, and the agency is advised to discuss the thresholds for its region with market and industry experts. It is important to remember, however, that not all possible economy-specific situations may fit into the regional thresholds.

The *second objective* of the validation is to ensure comparability of prices. The purpose of the price comparison is to compare comparable products—that is, "likes with likes." Loose item specifications or difficult survey areas may introduce product errors because the products whose prices are being compared may not be the comparable ones. For example, if a generic product specification—that is, a specification that does not define the brand to be priced—is used, it may embrace a wider range of products than was intended when the specification was drawn up. In this case, errors may be introduced into the estimated PPPs because the products differ significantly in their characteristics. Errors stemming from loose specifications or difficult survey areas cannot be attributed to the data of any particular economy because they are not errors at the economy level; the data are coherent at the subnational level.

The role of the RCAs and the Global Office is to ensure that collected prices are comparable by analyzing both the price data and metadata. Because product errors at the regional level may not be errors at the economy level, it is not always clear to the NCAs why edits are proposed. Thus even if an economy actively participated in detecting price and product errors, it is the responsibility of the program coordinators to guarantee the comparability of the final prices.

The points just made are mainly related to nonsampling errors. Sampling errors occur

because the prices on which the PPPs are based are collected from a sample of outlets rather than from all outlets. An important distinction between the two types of error is that sampling errors would disappear were prices collected by an enumeration of all outlets, but nonsampling errors would not because they would continue to occur regardless of the types of outlets or the number of observations. Sampling errors should be controlled before price collection through sample design; nonsampling errors should be handled before and during price collection through good survey design and management and after price collection through validation. ICP validation is mainly directed at nonsampling errors. The objective is to minimize, if not eliminate, the incidence of nonsampling errors among the regional price data after collection. This is achieved by editing and verification. As for sampling errors, the possible validation actions are limited. However, the RCAs should try to assess the plausibility and comparability of the subnational sample and outlet sample across the region.

Overall, prevention is preferable to correction. The incidence of nonsampling errors can be significantly reduced through good survey design and management. Price collections must be carefully planned, efficiently carried out, and properly supervised; item specifications must be sufficiently detailed to enable price collectors to identify products unambiguously in the outlets they visit; price collectors must be well trained, given clear instructions, and provided with price reporting forms that are user-friendly; fieldwork must be closely monitored to ensure that price collectors record the prices, quantities, and other data required; and staff engaged in processing and validating the prices must be properly trained and supervised. Validation complements good survey practice.

Intra-economy Validation



The first validation stage in the ICP validation process is intra-economy validation. Within this stage, it is verified that price collectors within the same economy have priced products

that match the item specifications and that the prices they have reported are correct. The relevancy of the survey frame—that is, the outlet and location sample—also needs to be verified.

The intra-economy validation can be carried out entirely by the NCA without the assistance or intervention of the RCA, although the RCA can provide a valuable second opinion about doubtful or marginal cases.

Intra-economy editing searches for discrepancies between the requested and reported prices and metadata and analyzes problematic values—first, among the individual prices an economy has collected for each item, and then among the average prices for these items. The editing and subsequent verification are the responsibility of the economy's NCA. These tasks are carried out without reference to the price data of other economies. When the NCAs complete the intra-economy validation, they provide the RCA with validated average prices for the items priced plus the importance indicator,

the coefficient of variation, the minimum/maximum (min/max) price ratio, and the number of price observations for each of the average prices reported. The NCAs are also encouraged to submit the price observations and related metadata in order to facilitate the intereconomy validation.

The intra-economy validation stage consists of three validation phases: initial data validation (A), statistical tests (B), and finalization of price data (C). During the first phase, the price data generated by price collectors are merged and the first checks are made. During the second phase, the data are analyzed more systematically using statistical means. During the final phase, the data are prepared for delivery to the RCA. Each phase and its underlying steps are described in the following sections.

Initial Data Validation



Step A1 Enter price and metadata into the validation software.

The first step in the validation process is to enter the individual prices and related metadata into the validation software. It is preferable that this be done by staff other than the price collectors. A strict division of labor between price collection and data entry ensures that the operations are completely objective. For the purposes of validation, data entry should be done as close in time to the price collection as possible—that is, while price collectors and outlet personnel still have a clear recollection of the circumstances prevailing when the prices were collected. Ideally, the individual prices would be entered and screened the day after their collection, which would allow field supervisors to catch and correct the mistakes of price collectors from the beginning.

Step A2 Check prices and metadata for errors and discrepancies.

The next step is to check the entries to ensure that (1) item codes are correct; (2) codes and the corresponding price observations have been entered in sequence; (3) there are

only numeric entries for reference quantity, quantity observed, and price observed, as well as for those fields, such as location of outlet or type of outlet, that use numeric codes; and (4) the quantity observed is in the same unit as the reference quantity. The NCAs should also check that the prices seem to be initially plausible and are in line with the general knowledge of the market. Box 15.1 gives two examples of quantity-related problems.

In addition, the following points should be considered:

- *Alternative outlets.* In most cases, price collectors have been given a list of items with detailed specifications and a list of outlets from which to collect the prices of the products. Price collectors are instructed to visit the outlets listed. If for some reason they are obliged to choose another outlet, the reasons should be explained, and detailed information should be provided about the outlet from which the prices were actually collected. The NCAs should ensure that the alternative outlets are good substitutes.
- *Alternative products.* Price collectors note in writing the prices of the alternative products

BOX 15.1

Errors of Quantity: Two Examples

Example 1. Often the price reported refers to a different unit of quantity than that requested and specified in the item list. The price reported may refer to the wrong weight—such as 250 or 500 grams instead of the required quantity unit of 1 kilogram—or it may be the price per egg instead of per dozen eggs, or per pint or gallon instead of per liter. However, price collectors are obliged to always report the actual unit of quantity, whether or not it corresponds with the quantity requested. Thus the use of a different quantity should not introduce an error as it will be recorded and known. The price collected can then be adjusted to convert it into the price for the quantity requested. If this is not feasible, the price may have to be deleted.

Example 2. Suppose a price collector is only able to obtain a price for a 1 kilogram package

of an item whose specification is 150–300 grams. Thus the size of the priced package falls well outside the predetermined range. In this case, an automatic proportional adjustment on the basis of the relative quantities may not be appropriate. In general, a bulk purchase cannot be treated as if it were no different than several smaller purchases adding up to the same quantity. In this case, a 1-kilogram package cannot be treated as if it were the same, from a consumer's point of view, as, say, four 250-gram packages. Thus the price for a kilogram package may have to be rejected because it is too difficult to make an appropriate quality adjustment for such a large difference between the reported quantity and the reference quantity.

on the paper forms provided, together with any accompanying notes, comments, or explanations. They are instructed to collect prices only for the items listed, if possible. If a product is not available exactly as specified in their list, price collectors can substitute another product within the same outlet whose characteristics are very similar to, even though not identical to, the item specification. The NCA should ensure that alternatively priced products are comparable with the rest of the priced products and with the item specification.

Step A3

Check that prices are plausible at the subnational level.

A large economy can be divided into different areas, with price collections carried out separately. It is essential that the NCAs ensure that the data received from different areas are correct and plausible based on knowledge of the region. Separate editing and validation

steps may have to be carried out for the different areas.

The relationship of the various areas to the economy as a whole resembles the one between the economies and the ICP region; even large differences in prices among different parts of the economy may be entirely appropriate. In any case, realistic national average prices have to be the end result.

Step A4

Check that prices are temporally plausible.

A temporal comparison analyzes possible inconsistencies in price data, which may be the result of price errors. Prices can be compared temporally either *between ICP rounds* or *between price collection periods within an ICP round*.

Because the intra-economy validation deals with the price observations and respective average prices at the item level, the most reasonable temporal validation is carried out between the

price collection periods within an ICP round. Temporal validation between ICP rounds can be conducted at the price level, but the following points have to be taken into account:

- Data at the item level may not be available for the previous round(s).
- If the item-level data are available, assurance is needed that the item definitions are comparable between the rounds—that is, the item parameters have not changed substantially and the reference units and quantities are the same.

A direct comparison of prices between ICP rounds may not be valuable because of real price level changes—inflation or deflation—within an economy. Thus without a consumer price index (CPI) adjustment, the consistency of the prices may be difficult to analyze. On the other hand, a direct comparison can be used to check whether a PPP price level change is in line with the respective CPI-observed change.

A temporal analysis between price collection months or quarters can be carried out without a CPI adjustment because prices are generally collected for identical item definitions. Because the average prices within a given year should be

relatively stable within an economy, large inconsistencies may be a sign of errors in the data. The possible reason for the variation should thus be investigated and a reasonable justification or correction provided.

A challenging aspect of temporal analysis is the difficulty in directly identifying the potential source of error. Potential sources can be either the previous or the current data set or, in the case of temporal analysis between ICP rounds, problematic CPI indexes. However, the main advantage of the temporal comparison is the opportunity to flag cases for further analysis.

Step A5

Check important items.

The official method for computation of the ICP 2011 round basic heading PPPs within a region is the weighted country product dummy (CPD-W); it weighs items to reflect the expected expenditure shares within the basic heading of the item. Because expenditure shares below the basic heading level are by definition not available, participating economies are required to classify all available items as either "important" or "less important." Validation of these indicators at the economy level is discussed in chapter 20 of this volume.

Statistical Tests

Initial data validation (A)

Statistical tests (B)

Finalization of price data (C)

After data have been added to the validation software and initially validated, four statistical tests should be conducted. The detailed price information and the statistical tests are read using two tables: the *average price table*, which shows average prices as well as the results of the min/max price

ratio test and the coefficients of variation for each item, and the *price observation table*, which gives the individual price observations, as well as the results of deviation and t-value tests. Table 15.1 is an example of an average price table, and table 15.2 is an example of a price observation table.

Table 15.1 Example of Average Price Table, ICP 2011

Product code	Product name	Importance	Number of price quotations	Average price	Min price	Max price	Min/max price ratio	Coefficient of variation
110111.10110	Rice, basmati, first degree	1	6	7.58	5.50	10.00	0.55	20.99
110111.10111	Rice, Uncle Ben's	1	8	14.35	6.00	25.00	0.24	37.97
110111.10112	Rice, white, 25% broken	2	4	4.61	3.20	8.50	0.38	48.72

Source: ICP, <http://icp.worldbank.org/>.

Note: Outliers are highlighted.

Table 15.2 Example of Price Observation Table, ICP 2011

Product code	Product name	Location code	Location name	Location type code	Outlet code	Outlet name
110111.10110	Rice, basmati, first degree	12	A	1	106	Market
110111.10110	Rice, basmati, first degree	11	B	1	119	Market
110111.10110	Rice, basmati, first degree	11	B	1	119	Market
110111.10110	Rice, basmati, first degree	11	B	1	119	Market
110111.10110	Rice, basmati, first degree	11	B	1	119	Market
110111.10110	Rice, basmati, first degree	11	B	1	119	Market
110111.10111	Rice, Uncle Ben's	12	A	1	101	Market
110111.10111	Rice, Uncle Ben's	12	A	1	101	Market
110111.10111	Rice, Uncle Ben's	12	A	1	101	Market
110111.10111	Rice, Uncle Ben's	12	A	1	101	Market
110111.10111	Rice, Uncle Ben's	11	B	1	119	Market
110111.10111	Rice, Uncle Ben's	11	B	1	14	Market
110111.10111	Rice, Uncle Ben's	12	B	1	14	Market
110111.10111	Rice, Uncle Ben's	12	A	1	56	Market
110111.10112	Rice, white, 25% broken	11	B	1	1	Market
110111.10112	Rice, white, 25% broken	11	B	1	1	Market
110111.10112	Rice, white, 25% broken	11	B	1	1	Market
110111.10112	Rice, white, 25% broken	12	B	1	101	Market

Source: ICP, <http://icp.worldbank.org/>.

Statistical tests are carried out for the prices that have been converted to a reference quantity—that is, to the same quantity or volume such as 1 kilogram or 1 liter. Box 15.2 discusses the suggested limits for the statistical tests, and boxes 15.3 and 15.4 present guidance on the editing of the price data based on the results of the statistical tests.

Step B1

Check average price table for problematic values using min/max price ratio test.

The first test is the min/max price ratio test, which checks the ratio of the minimum reference quantity price observed for the product to the maximum reference quantity price observed

BOX 15.2

Suggested Limits for Statistical Tests

For each statistical test introduced in this section, a proposal for acceptance limits is given. However, these limits are suggested as "default values," and they may have to be modified, depending on the situation. Each basic heading or aggregate may have its "own" limits or critical values. For example,

for the coefficient of variation test, limits on the prices of new cars have to be tighter than those on services, whose prices tend to vary more substantially. Moreover, it may be necessary to apply different limits for different countries if, for example, it is known that the subnational price differences are high.

Table 15.2 (Continued)

Outlet type code	Obs. date (mm-dd-yyyy)	Obs. qty.	Obs. price	Converted price	Deviation	T- value
1	5-26-2010	1	9.00	9.00	18.7	0.9
1	5-5-2010	1	8.00	8.00	5.5	0.3
1	5-19-2010	1	10.00	10.00	31.9	1.5
1	6-1-2010	1	6.00	6.00	-20.9	-1.0
1	6-1-2010	1	7.00	7.00	-7.7	-0.4
1	6-11-2010	1	5.50	5.50	-27.5	-1.3
1	6-9-2010	2	50.00	25.00	74.2	2.0
1	6-9-2010	3	40.00	13.33	-7.1	-0.2
1	6-10-2010	2	23.00	11.50	-19.9	-0.5
1	6-10-2010	1	20.00	20.00	39.3	1.0
1	5-26-2010	2	22.00	11.00	-23.4	-0.6
1	5-27-2010	1	15.00	15.00	4.5	0.1
2	5-28-2010	1	13.00	13.00	-9.4	-0.2
1	3-3-2010	0.5	3.00	6.00	-58.2	-1.5
1	3-3-2010	2	6.50	3.25	-29.5	-0.6
1	6-4-2010	1	3.50	3.50	-24.1	-0.5
1	6-11-2010	1	8.50	8.50	84.3	1.7
1	6-11-2010	1	3.20	3.20	-30.6	-0.6

BOX 15.3

Editing Price Data Based on the Results of Statistical Tests

Editing for product errors and price errors involves identifying prices that have high values—that is, prices whose values are determined to be either too high or too low vis-à-vis the average according to given criteria. The price may score a value for a given test that exceeds a predetermined limit, or its value may fall outside some prespecified range of acceptable values. Both are standard ways of detecting errors in survey data, and both are employed by the ICP. Prices with problematic values are not necessarily wrong. *But the fact that their values are considered to be outside of a prespecified range suggests that they could be wrong—they are possible errors and need to be investigated.* It is not ICP

practice to reject outright prices with problematic values. Rather, it is first established whether they are genuine observations. Once this is known, it can be decided how to deal with them. Prices with high values that are found to be wrong are errors and should be corrected or dropped, whereas prices with problematic values that are shown to be accurate observations are "outliers" and should be retained if they are part of the population defined by the rest of the price observations. In practice, it is not unusual for outliers that meet this criterion to be "corrected"—that is, discarded or replaced by an imputed value—in order to remove the "noise" they introduce into the data set.

BOX 15.4

Accounting for a High but Justified Coefficient of Variation (CV)

A high or even extreme CV may stem from reasons other than a straightforward product or price error. The price of the product may vary greatly among different types of outlets, or the product may not have been priced consistently across outlets because either the product specification is too broad or it has been interpreted differently by different price collectors. If the price observations are correct and a comparable product has been priced across outlets, the price variation arising from different outlet types is an economic fact of life. The average price should therefore be retained and the reason for the variation explained to the regional coordinating agency (RCA). It is possible that the mix of outlet types selected for the survey does not reflect the distribution profile of the product in question. This

possibility should therefore be investigated and, if necessary, the mix adjusted as appropriate by suppressing prices from those types of outlet that are overrepresented or by duplicating the prices from those types of outlets that are underrepresented. Items with average prices whose variation is caused by specifications that are too broad or by inconsistent pricing across outlets should be deleted, unless they are important and the basic heading does not have a sufficient number of items. In the latter case, the items and their average prices should be retained provisionally. This information should be noted in relation to the item. Later, it can be decided with the RCA whether the products should be dropped, retained, or split on the basis of what other national coordinating agencies in the region have reported.

for the product. The formula for calculating the min/max price ratio is

$$\text{min/max price ratio} = \text{min price}/\text{max price}.$$

The suggested limit for the min/max price ratio test is 0.5 or below. An item with a min/max price ratio below 0.5 fails the test and will be flagged in the average price table as having a high value that needs to be verified. The price data of all products having this flag should be analyzed and corrected or approved.

Step B2

Check average price table for problematic values using coefficient of variation test.

The second test is the coefficient of variation (CV) test, which checks the standard deviation for the product, expressed as the ratio of the average reference quantity price for the product to the standard deviation. The formula for calculating the CV is

$$\text{coefficient of variation} = (\text{standard deviation}/\text{average price}) \times 100.$$

The suggested limit for the CV test is 39 percent or below. An average price of the item with a CV greater than 39 percent fails the test and will be flagged in the average price table as having a high value that requires verification. The price data of all items having this flag should be analyzed and corrected or approved.

Step B3

Check price observation table for problematic values using deviation test.

The third test is the deviation test, which measures the deviation of the reference quantity price for a price observation from the average reference quantity price for the product. The formula for calculating the deviation is

$$\text{deviation} = [(\text{price} - \text{average price})/\text{average price}] \times 100.$$

The suggested deviation limit is 50 percent or below. A price observation with a deviation higher than this fails the test and will be

flagged in the price observation table as having a problematic value that needs to be checked. All prices having this flag should be analyzed and corrected or approved.

Step B4

Check price observation table for problematic values using t-value test.

The final test is the t-value test, which checks the ratio of the deviation of the reference quantity price for a price observation from the average reference quantity price for the product to the standard deviation of the

product. The formula for calculating the t-value is

$$t\text{-value} = [(price - average\ price)/standard\ deviation].$$

The suggested limit for the t-value test is 2.0 or below. A price observation with a ratio greater than 2.0 fails the test and will be flagged in the price observation table as having a high value requiring investigation. All prices having this flag should be analyzed and corrected or approved.

Finalization of Data

Initial data validation (A)

Statistical tests (B)

Finalization of data (C)

Step C1

Confirm that price data and metadata have been intra-economy validated.

Before the data and metadata are submitted, it should be confirmed that steps A1 to B4 were conducted successfully and that the data were intra-economy validated.

Step C2

Submit price data to the regional coordinating agency.

After the data are confirmed to be intra-economy validated, they can be sent to the RCA. The data and metadata have to be in a correct, prespecified format.

Intereconomy Validation

Intra-economy validation

Intereconomy validation

Global validation

The intereconomy validation stage starts as soon as a sufficient number of economies within the region have submitted intra-economy validated price data. Intereconomy validation is a collective process involving the RCA and a group of economies, and it is designed to establish that price collectors in different economies have priced products that are comparable between economies—in other words, they have all interpreted the item specifications in the same way—and that the prices they have reported are correct. Intereconomy editing looks for problematic values among the average prices as well as discrepancies in the metadata that the region's NCAs have reported to the RCA for the same items within a basic heading. For this step, the average prices, which are expressed in national currencies, have to be converted to a common currency. After being

converted, the average prices of each economy are checked against the average prices of the other economies in the region. This step cannot be carried out effectively without the active participation of the RCA and NCAs, which must agree to share their average prices.

The main tools for the intereconomy validation are the Quaranta table, named after Vincenzo Quaranta, who first proposed it for use in the European PPP program in 1990, and the Dikhanov table, named after Yuri Dikhanov, who first proposed it for use during the ICP 2005 round (for a detailed explanation of these tables, see chapter 14). Both validation tables are prepared by the RCAs, but the tables should be circulated to all economies within a region. The NCAs are expected to make their own assessments of their data relative to those of other economies independently of the RCA. In other

BOX 15.5

Editing of Price Data Based on Analysis of Quaranta and Dikhanov Tables

Editing a basic heading with a Quaranta table or an aggregate with a Dikhanov table entails identifying average prices with low or high values or, more precisely, the XR-ratios or PPP-ratios with low or high values. Average prices underlying these ratios flagged as deviant values are only possible errors. Regardless of how strictly the criteria used to identify these errors are defined, they are not necessarily errors by definition and therefore cannot be automatically removed. They have to be referred back for verification to the national coordinating agencies (NCAs) reporting them. The NCAs are required to investigate the average prices returned to them as possible errors and to confirm

whether they correct or incorrect. When prices are found to be incorrect, the NCAs are expected to correct them; otherwise, prices are suppressed. But if they are found to be correct, they are outliers, and a possibly difficult decision must be made about whether to keep them, to replace them with an imputed value, or to drop them. Some of the deviations, even larger ones, can be legitimate. Individual economies may have particular pricing policies, such as low fuel prices in some of the oil-producing countries. Such prices may be flagged as problematic values, but they would not be incorrect, and it would be wrong to remove them despite the "noise" they may introduce into the data set.

words, the NCAs are expected to be proactive during the intereconomy validation. The RCA may refer doubtful prices back to the economies for clarification or further investigation. The economies may then revise or correct their data, or seek to justify the data as reflecting the market situation in their economy. Intereconomy validation can involve several iterations or rounds before being completed, as explained at the outset of this chapter. After each round, as incorrect or incomparable prices are removed or corrected, the PPPs will become more reliable. Box 15.5 describes the editing of price data based on analysis of Quaranta and Dikhanov tables.

The intereconomy validation stage consists of six validation phases. The first phase, *initial data validation*, is a quick validation phase before any

in-depth validation to reveal any problems with the data that would make any deeper data analysis impossible by blurring the validation tables. The second phase is *validation of metadata*, which seeks to identify possible large deviations within requested and reported data and any other comparability issues that may not be visible as outlier indexes. *Validation at basic heading level* is the main phase of the intereconomy validation, and most of the edits are carried out within that phase. *Validation at aggregated level* takes basic heading average prices to a wider context. *Temporal analysis* checks the consistency of PPPs or price level indexes (PLIs) with regard to a previous ICP round or previous quarters within an ICP round. *Finalization of data* concludes the intereconomy validation phase.

Initial Data Validation

Initial data validation (D)

Validation of metadata (E)

Validation at BH level (F)

When the data received from the NCAs are compared for the first time, a number of severe errors are usually evident that make any in-depth

analysis impossible because the indexes in the validation tables are blurred by the potential extreme problems with the data. The most

common extreme errors are quantity-related or simple typing mistakes that may not be visible at the economy level but become clear when data for different economies are compared. Because the focus is only on the extreme problems with the data, one initial validation round should be enough to conclude this validation phase.

Step D1

Enter price and metadata into the validation software.

Intereconomy validation begins with the RCA entering the average prices reported by the NCAs into the validation software. The first problems with the data are normally encountered during this step, and they stem from incorrect data formats.

Step D2

Calculate initial validation tables.

After the prices have been entered, it is possible to produce an initial Quaranta table. It is recommended that the Quaranta table be used instead of the Dikhanov table in this validation phase because at the early phases of data validation, the BH PPPs and the respective item

PPP-ratios and country product dummy (CPD) residuals are not reliable (they are based on invalidated data). The Dikhanov table focuses on residuals using a color scheme, whereas the Quaranta table gives equal weight to the exchange rate ratios of price (XR-ratios) and the PPP price ratios (PPP-ratios).

Step D3

Check initial validation tables for extreme XR-ratios.

XR-ratios are standardized price ratios based on exchange rate-converted average prices. The ratios are thus a proxy measure that allows the average prices of a product to be compared across economies. Each of these ratios refers to a particular item in a particular economy. An extreme XR-ratio means that the national average price for the product in question is very much out of line when compared with the prices of the same product in other economies when they are all converted into the common numéraire currency using the exchange rate.

The focus should be only on very high or very low XR-ratios—that is, values approximately below 40 or over 300.

Validation of Metadata

Initial data validation (D)

Validation of metadata (E)

Validation at BH level (F)

In addition to the validation based on the validation tables, the metadata underlying the average prices should be checked separately against large deviations between requested and priced products. For example, if an economy has systematically priced products with larger quantities than are requested for a basic heading, the result may be bias that is not visible at the BH-level indexes. However, references must also be made to the metadata during the following validation phases when potential outlier prices are investigated.

Step E1

Check metadata for large deviations between requested and priced products.

Small deviations from the metadata should be allowed, but the deviation should be considered when a product is no longer the one priced

by the other economies. The common rules for the deviation permitted should be set while taking into account that these rules can differ across basic headings. Box 15.6 gives an example of the dialogue associated with edits based on the metadata validation.

Step E2

Check metadata to ensure that survey frames are comparable across economies.

Survey frames established by the national coordinating agencies follow the typical national purchasing patterns. However, errors or very different purchasing patterns may result in data that cannot be compared. For example, data from an economy that prices products solely at department stores in its capital city may not be comparable with those from other economies that collect

BOX 15.6

Example of Validation Dialogue for Men's Tennis Socks

Regional coordinating agency: The item specification asks for multipacks, but you have priced both single packs and multipacks. This results in a very high coefficient of

variation. Please remove the incorrect package sizes.

National coordinating agency: OK, incorrect package types deleted.

prices from a balanced mix of outlets and locations. RCAs should try to assess the comparability of national survey frames to the

greatest extent possible while understanding that, in practice, the means for this may be limited.

Validation at Basic Heading Level

Initial data validation (D)

Validation of metadata (E)

Validation at BH level (F)

Validation carried out at the basic heading level is the main phase of the intereconomy validation.

Step F1

Calculate validation tables.

After the data have been cleansed of extreme errors, the first "real" validation tables can be calculated. For this purpose, either the Quaranta or the Dikhanov table can be used.¹

Step F2

Check plausibility of BH price level indexes.

PLIs provide a comparison of the economies' price levels with respect to the regional average. If the PLI of an economy is higher than 100, its price level is expensive compared with the regional average and vice versa. PLIs are not intended to rank economies strictly. In fact, they only provide an indication of the order of magnitude of the price level in one economy in relation to others, particularly when economies are clustered around a very narrow range of outcomes.

During validation, PLIs can be used in two ways: (1) to check the consistency of relative price levels across basic headings for a single economy and (2) to check the ranking of economies for a single basic heading.

If for a single economy PLIs across basic headings vary considerably, it may be a sign of problems with the data for the respective "outlier" basic headings. However, the variation may also be justified if, for example, individual economies have particular pricing policies, such as low fuel prices in some of the oil-producing economies. Nevertheless, such a basic heading should be flagged for further analysis.

For the PLIs for a single basic heading, the ranking of economies tends to follow a certain pattern. If the price level of an economy is considerably more or less expensive than it usually is in the other economies within the group, it may again be a sign of problems with data for this basic heading.

Step F3

Check BH tables for high or problematic average coefficients of variation.

The average CV measures dispersion among all the PPP-ratios for a basic heading. In doing so, it measures the homogeneity of the price structures of the economies covered by the basic heading and the reliability of the PPPs calculated for the basic heading. The higher the coefficient's value, the less homogeneous are the price structures and the less reliable are the

PPPs. A value of over 39 percent is considered problematic.

During verification of high (low) or problematic values, priority should be given to basic headings with coefficient values greater than 39 percent, particularly if they have large expenditure weights. Basic headings with large expenditure weights will have a greater influence on the overall PPPs than basic headings with small expenditure weights. Only the Quaranta table shows expenditure weights—they are available at the time of the validation.² The value of the coefficient should fall as validation progresses, thereby providing a means of assessing the overall effectiveness of the validation process.

Step F4

Check BH tables for high or problematic economy coefficients of variation.

The economy CV measures dispersion among an economy's PPP-ratios for a basic heading. In other words, it measures the variation in an economy's price levels among the items it priced for the basic heading and the reliability of its PPP for the basic heading. The higher the coefficient's value, the less uniform are the economy's price levels and the less reliable are its PPPs. A value of over 39 percent is problematic.

During verification, priority should be given to basic headings for which the value of the economy CV is greater than 39 percent, particularly if the expenditure weight for the basic heading is large. The coefficient should decline in value as validation progresses. This allows an assessment of the effectiveness of the validation process.

The economy CV complements the item CV (see step F6) by introducing a different perspective to the same set of data. Focusing on economies rather than items can help to identify economies that have suspicious data.

Step F5

Check BH and item tables for important products.

The selection of important items should be verified across the economies within a region because the choice is based on estimated expenditure shares for the items within a basic heading. Validation of importance indicators at the

intereconomy stage is covered in chapter 14 in this volume.

Step F6

Check item tables for problematic item coefficients of variation.

The item CV could be said to be the most important of the CVs for validation purposes. It measures dispersion among the PPP-ratios for an item. It is an indicator of comparability and accuracy that addresses the following questions: Have the NCAs priced the same products or equivalent products? And have they priced them correctly? The higher the coefficient's value, the less uniform are the item's price levels and the more suspicious are the product's comparability and the accuracy of its pricing across economies. Such products are candidates for splitting or deletion, and the RCA should ensure that they are thoroughly investigated by the NCAs. During validation, priority should be given to products with a CV greater than 39 percent.

Step F7

Check item tables for problematic XR-ratios.

XR-ratios are standardized price ratios based on exchange rate-converted average prices. The ratios are thus a proxy measure that allows the average prices of a product to be compared across economies. Each of these ratios refers to a particular item in a particular economy. A high (low) XR-ratio means that the national average price for the item in question is high (low) compared with the prices of the same item in other economies when they are all converted into the common numéraire currency using the exchange rate.

However, it must be remembered that the principal reason for calculating PPPs is that when the prices of a given product are converted into a common currency unit using exchange rates, they are not in fact equal in all economies. The general level of prices tends to be systematically higher or lower in some economies than others. Thus a high or low exchange rate price for an individual item in one economy may be largely due to the fact that the general price level for that economy is high or low when exchange rates are used.

It may not signal any abnormality in that particular price.

For this reason, analysis of the XR-ratios facilitates the identification of problematic values among price ratios for an item at the beginning of the intereconomy validation when PPPs and PPP-ratios are likely to be unreliable because they are based on the average prices being validated. Initially, XR-ratios outside the range of 80–125 should be investigated during the first and second rounds of validation. In later rounds, when the PPP-ratios become more reliable, high (low) or even extreme values among the XR-ratios can be ignored. For this indicator, it is better to consult the Quaranta table because the series is clearly displayed. Box 15.7 addresses the issue of detected outliers.

Step F8

Check item tables for problematic PPP-ratios or CPD residuals.

PPP-ratios are standardized price ratios based on PPP-converted average prices. These ratios are the correct measure to use in comparing the average prices of a product across economies and the average prices of an economy across products, and they are the main validation indicator at the product level. It is thus the problematic values among these price ratios for an item that intereconomy validation seeks to identify and verify. CPD residuals in the

Dikhanov table are equal to the logarithms of the PPP-ratios in the Quaranta table.

High (low) PPP-ratios for the same product in different economies imply that the relative price of the item tends to vary considerably from economy to economy. Although plausible, this may signal that one or more of the prices of the items are incorrect. Ratios outside the range of 80–125 should be investigated, as should residuals smaller than -0.25 or higher than 0.25 . Boxes 15.8 and 15.9 provide examples of the dialogue associated with edits based on high PPP-ratios.

Step F9

Check item tables for problematic price observation coefficients of variation.

The price observation CV measures variation in the price observations on which the average price reported for an item by an economy is based. It is taken directly from the average price table, and it is used to identify problematic values among average prices during intra-economy validation when average prices with a CV of over 39 percent are considered problematic. Should the CV remain over 39 percent after intra-economy validation, the NCA may need to re-edit the underlying price observations if there are problematic values among the item's PPP-ratios or if the item CV is over 39 percent. Box 15.10 gives an example of the dialogue associated with edits based on high price observation CVs.

BOX 15.7

Outliers

The disturbance created by an outlier can affect not only the PPP for the economy reporting the outlier but also the PPPs for other countries in the regional comparison. In such cases, replacing the outlier with an imputed value or suppressing it is an option to be considered. If within the context of a basic heading the outlying average price refers to a product that is particularly

important for the reporting economy, deleting it may not be justified, but imputing a value may be. However, if the average price refers to a product that is less important or not important, removing it is probably warranted. Whatever action is taken, it has to be decided jointly by the economy's national coordinating agency and the regional coordinating agency on a case-by-case basis.

BOX 15.8

Example of Validation Dialogue for a Whiskey Product

Regional coordinating agency: This item is relatively very expensive because the XR-ratio is 207 and the PPP-ratio is 168. And you are the most expensive economy within the region, which is not normally the case. Please check that the correct products have been priced and that the quantity is correctly reported. Please also

note that within the previous quarter the average price for this item was in line with the other items within this basic heading.

National coordinating agency: Sorry, we priced 12-year-old whiskey, which is not according to the product specification. The prices are deleted.

BOX 15.9

Example of Validation Dialogue for Hairdressing Services

Regional coordinating agency: The PPP-ratio is high (160), and the average price for this dry haircut item is higher than for the next wet haircut item. Basically, this should not be possible. Please check. Also, if all services are offered typically

only with wash, you should price only the next item.

National coordinating agency: Only haircuts with wash are offered in our country. Prices are moved to the next item.

BOX 15.10

Example of Validation Dialogue for Loose Mushrooms

Regional coordinating agency: The coefficient of variation for loose mushrooms seems unusually high. In particular, price observations 5–10 should be checked again. Are you sure these prices are for the right product and quantity and that they were collected in representative outlets? In addition, observation 5 is marked as "sold loose." Are you sure your price collector weighed the product and gave the price for 500 grams?

Possible answers:

- The price collector may have priced a special offer by mistake. We have deleted the price.
- The price was collected from an unrepresentative small neighborhood shop. We have deleted the price.
- We have no explanation for the product. Our market specialists considered it exceptional: it has been deleted as an outlier.

Validation at Aggregate Level



Dikhanov tables can be compiled for a group of basic headings constituting an aggregate. Validation at the aggregate level puts the editing and verification of average prices in a broader context. In other words, to ensure consistency between the average prices, an analysis within the basic heading as well as within a larger set of products is conducted.

Editing at the aggregate level enables inconsistencies to be identified that would not be found by editing solely at the basic heading level. The validation steps carried out at the BH level and introduced in the previous section should also be conducted at the aggregate levels. For example, if for the basic heading "alcoholic beverages" an economy priced all the beverages in quarts instead of liters as specified, its price ratios would be consistent within the basic heading, but they would not be consistent with the economy's price ratios

in other basic headings. Such errors are identified by editing across basic headings. In this respect, it is useful to compile Dikhanov tables at different levels of aggregation. For example, basic headings covering food items could be first checked in a Dikhanov table covering food and nonalcoholic beverages and subsequently in a Dikhanov table covering the entire household final consumption expenditure.

Step G1 Calculate the Dikhanov tables.

This calculation should be carried out for different levels of aggregation.

Step G2 Conduct steps F2 to F9.

See previous section. Box 15.11 is an example of validation at the aggregate level.

BOX 15.11

Example of Validation at the Aggregate Level

Suppose that the PPP for a fruit aggregate for economy A is about 40 percent higher than the PPP for a vegetable aggregate, whereas for the other economies the two sets of PPPs are

of a similar order of magnitude. This result may reflect reality, or it may indicate a possible issue with the prices collected and recorded in economy A. That possibility must be verified.

Temporal Analysis



Temporal validation at the regional level allows comparisons to be made using PLIs or PPPs, in addition to item-level average price comparisons. Comparisons of PLIs or PPPs are more robust over time because they are not comparing directly the average prices for certain items whose availability is not ensured

or whose item specifications may have changed between the two comparison points. Nevertheless, the basket of items as well as participating economies may have an impact on the results that is not related to the actual price level changes within individual economies.

In the temporal comparison, the source of potential errors may be difficult to determine. It can derive either from the previous or current data set or from the CPI indexes. In any case, the idea of a temporal comparison is to flag cases for further analysis.



As during the intra-economy validation, two types of temporal comparisons can be carried out: one between the ICP rounds or one between months or quarters within the ICP round.

In most cases, tables for these comparisons need to be created manually using the available

BH and aggregate PPPs or PLIs. For the comparison between ICP rounds, the CPI adjustment has to be applied, or, alternatively, the magnitude of the PPP price level change can be compared with that of the CPI price level change.



If differences in item average prices, PLIs, or PPPs are observed, they should be investigated. If errors are not found, a reasonable justification for the differences should be provided.

Finalization of Data



Intereconomy validation is an iterative process. It can commence before all economies participating in the regional comparison have supplied their average prices. After each iteration or verification round, the RCA will change the region's price database in line with the findings reported by the NCAs of economies covered in the round, add to the database the prices of economies joining the validation process, and produce new Quaranta and Dikhanov tables. These tables will identify new problematic values as a result of the changes introduced by the RCA, and those values will have to be investigated by the NCAs. Gradually, after a number of rounds of verification in which the prices of all economies participating in the comparison have been included in the database, a convergence will occur, and the return on further rounds of

verification will be deemed marginal by the NCAs and the RCA and therefore not worth pursuing. Consequently, the intereconomy validation will be considered finalized.



In signing off on the validation process, the NCAs and the RCA formally attest that the price data were intereconomy validated.



After the price data are confirmed to be intereconomy validated, they are sent to the Global Office. The data and metadata have to be in the correct, prespecified format.

Global Validation



After the data are validated at the economy and regional levels, the final validation stage is the global validation in which the NCAs,

RCAs, and Global Office verify the prices for the GCL items, regional PPPs, and subsequent global PPPs. The objective is to ensure

that prices collected for the GCL items across the regions are comparable and that the linked global PPPs are plausible in wider terms.

The global validation stage includes three validation phases: (1) analysis of the regional validation tables, (2) validation of the GCL items, and (3) validation of the global PPPs.

Analysis of Regional Validation Tables



The global validation stage should start with analysis of the regional price data and respective validation tables in order to identify any problems at the regional level that could lead to comparison problems at the global level. The focus should be on the price variation at the economy, basic heading, and item levels, as well as on the classification of important items. As for the latter, the shares of important and

less important items should be analyzed, and the PPPs and PLIs calculated using different methods (CPD and CPD-W) should be compared in order to study the impact of the important items. Especially critical is ensuring that the classification of important items is comparable across regions in view of the fact that the CPD-W is the agreed-on BH PPP calculation method for the ICP 2011 round.

Validation of Global Core List Items



Central to the ICP 2011 round are the GCL items that are used to link the regional PPPs to form a global set of PPPs. It is essential that the price data for GCL items be validated thoroughly at the global level from both the "regional" and "global" perspectives.

Regional Perspective

An assessment is needed of how well the GCL items reflect the regional results as a whole. This assessment can be carried out by analyzing the regional validation tables but, most important, by comparing PLIs and PPPs based on (1) a full set of regional items, (2) only GCL items, or (3) regional items excluding GCL items. GCL items that are obvious outliers may be deleted in the global validation process if necessary after review and consultation with the RCAs and the economies concerned.

Global Perspective

To gain the global perspective, the price data for the GCL items should be validated across

regions by calculating PPPs and PLIs for all ICP economies. This step is a challenge because the large number of economies makes it difficult to apply the conventional Quaranta or Dikhanov table.

A top-down validation approach can be used by starting the validation at the class and BH levels and moving to more detailed levels if problematic cases are encountered. This approach consists of the following steps:

- Calculate PPPs and PLIs for all economies based on the GCL items.
- Compile BH and aggregate matrixes for PLIs and economy coefficients of variation.
- Flag PLI and CV matrixes for problematic cases.
- Analyze the flagged cases using country diagnostic reports and Quaranta tables.

Table 15.3 is an example of a PLI matrix, and table 15.4 is an example of a CV matrix. In the PLI matrix, potentially problematic cases are identified by searching for discrepancies

Table 15.3 Example of Price Level Index Matrix, ICP 2011

	Economy 1	Economy 2	Economy 3
Group	94.3	88.9	119.4
Class 1	97.9	73.2	139.6
BH 1	145.4	84.4	81.5
BH 2	98.2	92.5	110.1
BH 3	111.9	46.3	192.9
BH 4	81.6	99.1	123.6
Class 2	86.2	135.3	85.8
BH 5	70.0	126.9	112.5
BH 6	112.8	116.8	75.9
BH 7	47.1	160.5	132.1
BH 8	67.1	218.5	68.2

Source: ICP, <http://icp.worldbank.org/>.

Note: BH = basic heading.

(high or low values) between the economy's PLI for a class (e.g., bread and cereals) and the PLIs for the basic headings (e.g., rice) composing the class. Thus it is assumed that basic headings under a class would have similar price levels for an economy. The limits for identifying discrepancies can vary based on the validation round and the composition of the economy. Typically, a first validation round has wider highlight limits than subsequent rounds. Furthermore, if the group of economies being analyzed has very similar price levels, the limit can be tighter than in the case of economies with very different price levels. For example,

Table 15.4 Example of Coefficient of Variation Matrix, ICP 2011

	Economy 1	Economy 2	Economy 3
Group	10.6	8.8	13.0
Class 1	10.8	9.1	12.8
BH 1	41.3	3.5	42.0
BH 2	20.0	21.1	27.0
BH 3	2.5	8.8	1.8
BH 4	6.5	10.7	13.1
Class 2	10.5	8.4	13.2
BH 5	20.6	7.3	19.6
BH 6	1.2	1.1	3.5
BH 7	41.0	15.0	40.1
BH 8	12.0	42.0	11.1

Source: ICP, <http://icp.worldbank.org/>.

Note: BH = basic heading.

economy 1 in table 15.3 has a PLI of 145.4 for the first basic heading, which is considerably higher than the PLI value for the class (97.9), and thus it is highlighted in red for further analysis.

In the CV matrix, the cases requiring attention can be identified by searching for values higher than the pre-agreed limit, typically over 40. As noted earlier in the section on inter-economy validation, the average CV measures dispersion among all the PPP-ratios for a basic heading, or for a higher aggregate such as class. In table 15.4, five problematic cases require further analysis.

Validation of Global PPPs



The global PPP validation phase comes after the price data for both purely regional and GCL

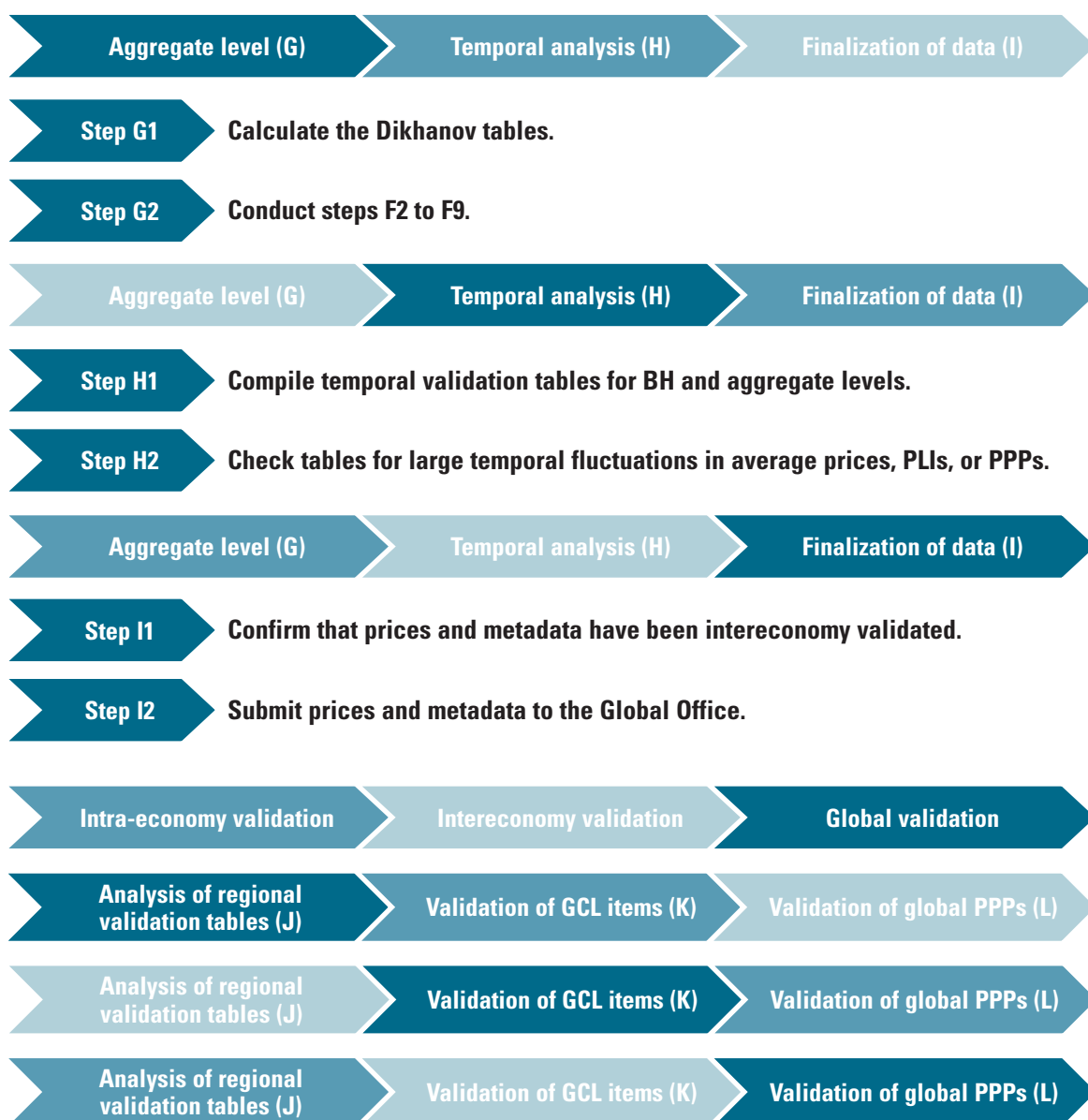
items have been validated. The steps to be carried out are described in chapter 22 of this volume.

Annex

Summary of Validation Stages, Phases, and Steps







NOTES

1. This statement leaves open the question of whether intereconomy validation should commence with the Quaranta table or the Dikhanov table. Some analysts prefer to start with the Quaranta table and consult the Dikhanov table after several rounds of verification have been conducted and the

PPPs are more reliable. Others prefer to begin with the Dikhanov table, using it to identify economies and items that require investigation, and subsequently organize the investigation around the Quaranta table.

2. The weights in the Quaranta table are only presented; they are not used in the calculation of PPPs.