



**ECONOMIC COMMISSION FOR
EUROPE and
CONFERENCE OF EUROPEAN
STATISTICIANS**

**EUROPEAN FREE TRADE ASSOCIATION
STATE STATISTICAL COMMITTEE OF
UKRAINE**

Joint EFTA/UNECE/SSCU Seminar “Economic Globalisation: A Challenge For Official Statistics”

Kiev, 3-6 July 2007
Session 4 of the provisional agenda

COMPILATION OF CONSUMER PRICE INDICES (CPIs)

Prepared by Carsten Hansen, UNECE

INTRODUCTION

1. The activities involved in the compilation of a CPI can be divided in the following areas of work:

- (a) Determining the purpose(s) and conceptual basis of the CPI
- (b) Deciding on the index coverage and classification structure
- (c) Compilation of expenditure weights
- (d) Designing the sample of outlets and products
- (e) Price collection
- (f) Calculation of the index
- (g) The treatment of missing prices and replacement products
- (h) Adjusting for quality changes
- (i) Data editing
- (j) Dissemination

2. The paper describes briefly these areas and aims to point only on the most important issues within the various activities. The paper relies heavily on the international *CPI Manual*, published in 2004¹, to which interested readers are referred for further details and explanations.

I. DETERMINING THE PURPOSE(S) AND CONCEPTUAL BASIS OF THE CPI

CPI used for different purposes

3. A consumer price index (CPI) measures the average rate of change of prices of goods and services acquired by the households for consumption. Most statistical offices compile only one CPI, which serves different purposes:

- as a measure of the general rate of consumer price inflation
- as a measure of changes in the cost of living
- deflation of national accounts series
- indexation of wages and pensions and the like
- indexation of private contracts

4. It is useful to discuss and decide on the main purpose(s) of the index, as this should serve as a basis for subsequent decisions regarding the compilation of the index.

Inflation measure

5. In the theory on price indices there are two main types of CPIs, *fixed basket indices*, and *cost of living indices*. The fixed basket index tries to measure the average price change of a representative basket of goods and services that is kept constant over time. This type of an index is sometimes referred to as a *pure price index*, a *cost of goods index (COGI)* or an *inflation index*.

Cost of living measure

6. A cost of living index (COLI) tries to measure the change in the cost of maintaining a given standard of living over time. The quantities are allowed to differ between the periods compared as a result of households' substitution behaviour. In the fixed basket approach by definition the quantities are kept constant.

7. In practice, however, the CPI has to be calculated on the basis of a fixed basket or fixed weights approach. Whether the primary purpose is to measure the inflation or the cost of living may primarily influence the coverage of certain goods or services and the treatment of owner-occupied housing.

¹ ILO, IMF, UNECE, OECD, Eurostat and the World Bank: Consumer Price Index Manual: Theory and Practice. ILO 2004, Switzerland. The manual is available in English from ILOs website: <http://www.ilo.org/public/english/bureau/stat/guides/cpi/index.htm>

The price concept 8. The CPI should be based on *purchaser prices* – the prices that the household actually pay for goods and services, including taxes less subsidies.

II. DECIDING ON THE INDEX COVERAGE AND CLASSIFICATION STRUCTURE

Population coverage 9. For the CPI to be representative, the weights of the index should be defined on the basis of the consumption of the whole reference population including all households. More narrowly defined CPIs may be compiled for special purposes, for example a CPI for low-income households or wage and salary workers types of households, where the weights would have to be derived from the expenditure data for this limited groups of households.

Geographical coverage 10. Ideally, the CPI is an indicator for the entire country. It is therefore important that the compilation basis is representative for the whole country. In principle, the expenditure weights and the prices collected should refer to the same population and geographical area. However, due to resource constraints, price collection is often limited to urban areas the price changes of which are assumed to reflect those in surrounding smaller areas.

'National' and 'domestic' concepts 11. In the definition of population and geographical coverage, countries tend to follow one of two options:

- The *national concept*, which covers the consumption expenditure of resident households, whether made in the country or abroad
- the *domestic concept*, which covers the consumption expenditure inside the country, whether made by resident or foreign households

12. It is up to the statistical office to decide on the coverage, depending on the purpose of the CPI. The Harmonized consumer price indices (HICPs) of the EU member countries, for example, follow the domestic concept as this is seen to be the more relevant for monitoring inflation.

Coverage of goods and services 13. The goods and services consumed by the households can in principle be acquired in four ways:

- (a) Purchase in monetary transactions
- (b) From own production
- (c) As payment in kind
- (d) As transfers or gifts from other economic units

14. The broadest possible scope for goods and services would cover all four categories. It would include all social transfers in kind in the form of education, health and housing and other goods or services provided free of charge or at nominal prices. The total acquisition of goods and services is equivalent to total actual consumption of households in the SNA.

15. For the CPI as a measure of inflation the more relevant would be to include only goods and services purchased in monetary transactions by the households, which excludes categories 2) – 3). Only monetary expenditures generate prices that can be observed for the CPI. The HICP is based on *households' final monetary consumption expenditure*.

Owner-occupiers and food produced for own consumption

16. However, many countries include the service of owner-occupied housing even if there is no market transaction involved. The 'price' is imputed by the rent of similar rented dwellings or estimated by user costs. Food produced for own consumption is sometime included in the weighting data for the CPI although no prices are imputed.

Classification – COICOP

17. The individual goods and services should be grouped and aggregated according to an international, or international comparable, classification. The one most common in use is *COICOP – Classification of individual consumption according to Purpose*. COICOP derives from the SNA and is used for the HICP, for example.

III. COMPILATION OF EXPENDITURE WEIGHTS

18. The expenditure weights for the CPI are used to average the detailed, elementary aggregate indices into indices at higher level of aggregation, up to the overall CPI itself.

The household budget survey

19. The principal data source for household consumption expenditures in most countries is a household budget survey (HBS). When using HBS data appropriate corrections should be made to adjust for systematic bias, for example underreporting of 'undesirable' products such as gambling, alcohol and tobacco. Other data sources may be used for such purposes.

Adjusting HBS data

20. It may also be necessary to adjust HBS data to bring them in line with the consumption concept required for the CPI. Some important areas are *owner-occupied housing*, where the imputed rents are usually not included in the HBS; the *consumption of foreign households* (tourists) in the country, which also is not included in the HBS, but will have to be included in the weighting data if the domestic concept is applied; and *insurances*, where the weights should be net of claims paid out to be consistent with national accounts principles. Finally, the weights should include also expenditures of *institutional households* to be consistent with national accounts.

National accounts

21. Many countries also use consumption data from the national accounts to form the CPI weights, or they are used in combination with the weights from the HBS. The national account data can be used to compile weights at an aggregate level, for example for some 70 main groups of households expenditures, which can then be further divided by use of HBS data. This approach ensures, at least at some level of aggregation, consistency between the CPI and the national accounts.

Purchases between households 22. The weights should reflect the consumption pattern of the household sector as a whole. Buying and selling *between* individual households cancel out and should not influence the CPI weights.

Second-hand products 23. Second-hand products, for example used cars, should enter into the weights calculation with the amount used to purchase the product minus the amount received from sale of the product.

IV. DESIGNING THE SAMPLE OF OUTLETS AND PRODUCTS

24. The sample should provide a representative basis for the calculation of the CPI. In the design of the sample the following issues should be considered.

Geographical and population coverage 25. The sample should be representative of the geographical area covered by the CPI. It should cover larger and smaller cities and rural areas. Differences in the price *level* between regions are irrelevant to the CPI as a measure of price *changes*. Due to resource constraints the price collection may be restricted to mainly urban areas.

Coverage of outlets 26. The sample should be representative of the various types of outlets, supermarkets, chain stores, specialized outlets etc. The selection of outlets should be based on their market share, or the relative turnover, of the outlets.

Probability sampling 27. Outlets can be selected from business registers or by use of other sources, i.e. other statistics, telephone books, Internet. The outlets can be selected in a stratified probability sample where the probability of being selected is proportionate to the relative market share of the outlets.

Cut-off sampling 28. However, in many instances it will not be possible to conduct probability sampling. Another option is to use Cut-off sampling where focus is to ensure that the most important outlets are included, while outlets with smaller market shares are left out.

Coverage of goods and services 29. Most elementary aggregates cover a large variety of individual goods and services of which usually only relative few are included in the sample. In the sampling of goods and services it therefore has to be taken into account that the price development of the selected items has to represent the average price development of all items in the elementary aggregate.

Resources and response burden 30. It is resource demanding to collect prices and maintain the sample over time. It may therefore be useful for the statistical office to analyse and optimize the sample of outlets and products. This may save resources and reduce the response burden. For an optimal allocation the price volatility and weight of each individual item and the costs of obtaining the price has to be taken into account.

31. A simple, first approach, would be to calculate how much each elementary aggregate index, on average, contributes to the monthly or annual changes of the overall CPI. For example, the 12-month rate of change of the overall CPI can be decomposed into the percentage contributions of the elementary aggregate indices. This calculation can be repeated each month for a period of a year or more, and the average relative contributions be calculated. The relative importance of the elementary aggregates can then be compared with the numbers of prices collected for the elementary aggregates. This simple approach would provide a rough guidance about where the number of collected price may be reduced or increased.

V. PRICE COLLECTION

32. Prices can be collected in various ways:

- (a) By mailed questionnaires to the outlets.
- (b) By price collectors visiting the outlets.
- (c) By use of price lists, catalogs and Internet.
- (d) By central price collection.

Mailed questionnaires

33. Mailed questionnaires are a relative cheap way of collecting prices. It is more appropriate for areas with standardised products with less replacement and quality changes. The respondents have to be reminded to include new items that appear on the market, and delete obsolete ones.

Price collectors

34. The use of price collectors is resource demanding, but may be needed to ensure the quality of the collected prices, including updating of the sample with new items. It is more appropriate for areas with many product replacements and quality changes, fresh food and clothing, for example.

Price lists and Internet

35. Price lists and the like and Internet can also be used as data sources for the CPI, providing useful information at relatively low costs.

Central price collection

36. Prices for specific items may be collected from a central source. For example the prices of many supply services (water, gas, electricity) may be collected from one source. Similarly, outlet chains may charge the same prices in which cases it may be possible to collect the prices from a central office.

VI. CALCULATION OF THE INDEX

CPI calculated in two stages

37. The CPI is calculated in two stages. In the first stage the *elementary aggregate price indices* are calculated for groups of relatively homogenous goods and services, on the basis of the collected prices. In the second stage indices at higher level of aggregation, up to the overall CPI, are calculated

by weighting together the elementary aggregate indices with their relative expenditure shares.

VI.1 Calculation of elementary indices

Formation of elementary aggregates

38. The elementary aggregate indices are the basic building blocks of the CPI. In the formation of elementary aggregates some key points should be noted:

- Elementary aggregates should consist of groups of goods or services that are as similar as possible, and preferably fairly homogeneous.
- They should consist of items expected to have similar price movements. The objective should be to minimize the dispersion of price movements within the aggregate.
- The elementary aggregates should be appropriate to serve as strata for sampling purposes.
- The items selected should be ones for which price movements are believed to be representative of all the products within the elementary aggregate.
- The number of items within each elementary aggregate should be large enough for the estimated price index to be statistically reliable.
- The object is to try to track the price of the same item over time for as long as it continues to be representative. The items selected should therefore be ones that are expected to remain on the market for some time, so that like can be compared with like and problems with replacements and quality changes be reduced.

39. The elementary indices can be calculated in various ways. The most common practice is to calculate the elementary indices without using explicit weights for the individual observations. Three index formulas are common in used:

Carli index 40. The Carli index is the unweighted arithmetic mean of the price ratios:

$$(i) \quad P_{0,t}^C = \frac{1}{n} \sum \left(\frac{P_t^i}{P_0^i} \right)$$

Dutot index 41. The Dutot index is the ratio of the unweighted arithmetic mean prices:

$$(ii) \quad P_{0,t}^D = \frac{\frac{1}{n} \sum P_t^i}{\frac{1}{n} \sum P_0^i} = \frac{\sum (P_t^i / P_0^i) \cdot P_0^i}{P_0^i}$$

Jevons index 42. The Jevons index is the unweighted geometric mean of the price ratios, identical to the ratio of the unweighted geometric mean prices:

$$(iii) \quad P_{0:t}^J = \prod \left(\frac{p_t^i}{p_0^i} \right)^{1/n} = \frac{\prod (p_t^i)^{1/n}}{\prod (p_0^i)^{1/n}}$$

43. The Carli and Jevons are independent of the price levels. The Dutot index depends on the price levels in the reference period.

44. The Dutot index requires prices to be averaged. Hence, it should only be applied for elementary aggregates consisting of homogenous products nominated in the same measurement unit.

45. Table 1 shows an example of an elementary aggregate calculated by the three index formulas.

Table 1. Calculation of elementary indices

	January	February	March	April	May	June	July
<i>Prices</i>							
Item A	6,00	6,00	7,00	6,00	6,00	6,00	6,60
Item B	7,00	7,00	6,00	7,00	7,00	7,20	7,70
Item C	2,00	3,00	4,00	5,00	2,00	3,00	2,20
Item D	5,00	5,00	5,00	4,00	5,00	5,00	5,50
Arithmetic mean	5,00	5,25	5,50	5,50	5,00	5,30	5,50
Geometric mean	4,53	5,01	5,38	5,38	4,53	5,05	4,98
<i>Monthly price ratios</i>							
Item A	1,00	1,00	1,17	0,86	1,00	1,00	1,10
Item B	1,00	1,00	0,86	1,17	1,00	1,03	1,07
Item C	1,00	1,50	1,33	1,25	0,40	1,50	0,73
Item D	1,00	1,00	1,00	0,80	1,25	1,00	1,10
Carli – arithmetic mean of price ratios							
Monthly index	100,0	112,5	108,9	101,9	91,3	113,2	100,1
Chained monthly index	100,0	112,5	122,5	124,8	113,9	128,9	129,0
Direct index	100,0	112,5	125,6	132,5	100,0	113,2	110,0
Dutot – ratio of arithmetic mean prices							
Monthly index	100,0	105,0	104,8	100,0	90,9	106,0	103,8
Chained monthly index	100,0	105,0	110,0	110,0	100,0	106,0	110,0
Direct index	100,0	105,0	110,0	110,0	100,0	106,0	110,0
Jevons – geometric mean of price ratios = ratio of geometric mean prices							
Monthly index	100,0	110,7	107,5	100,0	84,1	111,5	98,7
Chained monthly index	100,0	110,7	118,9	118,9	100,0	111,5	110,0
Direct index	100,0	110,7	118,9	118,9	100,0	111,5	110,0

Direct and chained indices 46. In a direct elementary index, the prices of the current month are compared directly with those of the price reference month (January). In a chain index, prices in each month are compared with those of the previous month, and the resulting short-term indices are multiplied, or chained, together to obtain the long-term index.

47. On the basis of the Table 1 calculations the following points may be noted:

- The chained Carli index is upward biased and should not be used. In May all prices have returned to the initial level, but the chained Carli shows an increase of almost 14%. In July, all prices have increased by 10% while the chained Carli shows an increase of 29%.
- Jevons and Dutot are transitive – the direct index is equal to the chained index. The Carli index is not transitive.
- Jevons is always below the Carli.
- Carli shows an increase in case of ‘price bouncing’, i.e. the prices change between items but are otherwise equal, as from March to April.

48. From a statistical viewpoint the Jevons index appear to have better properties; it satisfies more tests in the so-called *axiomatic approach* to index number theory.

49. From an economic interpretation Carli and Dutot are fixed basket indices that maintain the underlying quantities unchanged. The Jevons index allow for some degree of substitution.

50. It is up to the statistical office to decide whether to calculate the elementary indices as direct or chained indices. In practice the use of monthly chained indices seems to have some advantage. When a new or replacement item is included in a direct index, it will often be necessary to estimate the price of the item in the price reference period, which may be some time in the past. In a chain index, new items are linked into the index as part of the ongoing index calculation by including the item in the monthly index as soon as prices for two successive months are obtained.

*Calculation of
weighted elementary
indices*

51. If weighting information about the individual items is available this can be exploited by calculating the elementary aggregates using the weights. Thus, the elementary aggregates can be calculated as Laspeyres indices or as so-called geometric Laspeyres indices:

$$(iv) \quad P_{0,t}^{La} = \sum w_0^i \left(\frac{P_t^i}{P_0^i} \right), \quad \sum w_0^i = 1$$

$$(v) \quad P_{0,t}^{GLa} = \prod \left(\frac{P_t^i}{P_0^i} \right)^{w^i}, \quad \sum w^i = 1$$

52. Note that expenditure weights should not be attached to individual price observations in a Dutot Index, since this would conflict with the implicit weighting of the price changes in this index.

VI.2 Calculation of higher-level indices

53. The higher-level indices are calculated as the expenditure weighted arithmetic average of the elementary aggregate indices:

$$(vi) \quad P_{0:t} = \sum w_b^j \cdot P_{0:t}^j$$

54. $P_{0:t}$ is any higher-level index, w_b^j is the weight of elementary aggregate j that belongs to the higher-level index, and $P_{0:t}^j$ indicates the corresponding elementary indices. The weights refer to a period b , which in practice has to precede period 0, the price reference period.

Additivity 55. Equation (6) applies at each level of aggregation above the elementary aggregate level; the index is additive. This means that any higher-level index can be calculated in one step as the weighted arithmetic average of the elementary indices of which it consist, or by weighting together the indices at intermediate level, with the same result.

56. Note that equation (6) is not a Laspeyres index. For equation (6) to equal a Laspeyres index, the weights should refer to the same period as the prices.

Price-updating of weights 57. The statistical office has to decide if the weights should be re-referenced, or *price-updated*, from the weight reference period (b) to the price reference period (0), or be applied as they stand without price-updating. Weights are price-updated by multiplying the original period b expenditure shares by their elementary indices from period b to 0, and rescaling to sum to unity:

$$(vii) \quad w_{b(0)}^j = \frac{w_b^j P_{b:0}^j}{\sum w_b^j P_{b:0}^j}$$

58. By price-updating the underlying quantities are implicitly kept constant. Price-updating the weights from b to 0 means that the index will show the same rate of changes as if the weights had been applied from period b .

Lowe index 59. If the statistical office decides to price-update the weights, the resulting index will be a *fixed basket index*, a so-called *Lowe index*. From month to month it measures the value of the same (annual) basket of goods and service. If the primary aim is to compile a CPI measuring the price development of an actual, past fixed basket of goods and services, the weights should be price-updated. The index will provide a good estimate of the price development if quantities tend to remains constant, i.e. if there is no or not much substitution in response to changes in the relative prices.

Young index 60. The statistical office can decide instead to use the expenditure shares without price-updating. This corresponds to the calculation of a *fixed weight index*, a so-called *Young index*. Focus is that the weights should be as representative as possible for the average expenditure shares in the period for which the weights are applied. It does not measure the changing cost of any actual fixed basket. It will provide a good estimate of the price development if the expenditure shares roughly remain constant, i.e. if consumers tend to substitute in response to relative price changes.

Automatic price-updating 61. In all cases automatic price updating should be undertaken with care. When there is an inverse relation between movements of prices and quantities, price-updating on its own could produce perverse results. For example, the price of computers has been declining rapidly for years and price-updating of the expenditure share of computers would seriously underestimate the importance of computers. In practice, the expenditure share of computers might actually be rising because of a very rapid increase in quantities of computers purchased.

VI.3 Chaining and re-weighting

62. From time to time a new set of weights is applied for the CPI. The weights can be updated annually or with less frequency. The weights should be updated at least every five years. Irrespective of the frequency of weights updating the CPI based on the new weights should be chained onto the terminal value of the CPI based on the old set of weights. The chained index is calculated as:

$$\begin{aligned}
 P_{0:t} &= \sum w_b^j \cdot P_{0:k}^j \cdot \sum w_c^j \cdot P_{k:t}^j \\
 \text{(viii)} \quad &= P_{0:k} \cdot \sum w_c^j \cdot P_{k:t}^j \\
 &= P_{0:k} \cdot P_{k:t}
 \end{aligned}$$

63. There are several important features of a chained index:

- The chained index formula allows weights to be updated and facilitates the introduction of new items and removal of obsolete ones.
- To chain the old and the new series, an overlapping period (k) is needed where the index is calculated using both the old and new set of weights.
- A chained index may have two or more links. In each link the index is calculated as a fixed weight index as in equation (6). The chaining period is usually a month, often December.
- Chaining is intended to ensure that the individual indices on all levels show the correct development through time.
- Chaining leads to non-additivity. When the new series is chained onto the old as in equation (7), the higher-level indices cannot be

calculated as the weighted arithmetic averages of chained individual indices.

64. It is useful to take the opportunity of re-weighting and chaining to introduce other changes to reduce breaks in time series.

VII. TREATMENT OF MISSING PRICES

Missing prices are inevitable, and have to be dealt with

65. Missing prices occurs regularly since is not always possible to collect all prices for the goods and services in the sample. Prices will be missing because outlets are closed or have stopped selling a product, or products have disappeared from the market, temporarily or permanently. A special type of temporary unavailability occurs with seasonal items such as fruit, vegetables, clothing and package holidays.

66. In the case of missing observations one of four actions may be taken:

- Omit the item for which the price is missing so that a matched sample is maintained, even though the sample is depleted.
- Carry forward the last observed price.
- Impute the missing price by the average price change of the prices that are available in the elementary aggregate.
- Impute the missing price by the price change of a comparable item.

Omit the item

67. One should be careful to omit too many observations as this will lead to a depletion of the sample.

Carry forward

68. Carrying forward prices biases the rate of change of the CPI towards zero. Thus, prices should only be carried forward for a limited period of time. When the item reappears, the whole of the price change should be included in the index.

Imputation by group

69. Imputing by the average price change of those goods and services for which prices have been recorded assumes that the price of the missing item would have moved as the average price change of the available items.

Imputation by another item

70. In some cases it may be more appropriate to estimate the price change of the missing price by the price change of another similar item, which may be known or expected to move in similar way.

Overlapping prices

71. If possible it may be an advantage to obtain overlapping prices, i.e. prices for the same month for both the disappearing and the replacement item. In this case, it is possible to link the price series for the new item to the price series of the item it replaces. Linking with overlapping prices involves an implicit adjustment for the difference in quality between the two items, since it assumes that the relative prices of the new and old item reflect their

relative qualities. For perfect or nearly perfect markets, this may be a valid assumption, but for other markets it may not be a reasonable assumption.

72. More practical examples of how to deal with missing prices are provided in the CPI Manual, chapter 9.

VIII. ADJUSTING FOR QUALITY CHANGES

73. The CPI shall reflect only 'real' price changes. Thus, price changes that are due to quality changes as a matter of principle should not be included in the CPI. Quality adjustments can be undertaken in a variety of ways:

- (a) **Direct comparison:** The price of the new item is compared directly with the price of the old one. It is thus assumed that the two items are of comparable quality and the whole of the price change is included in the index.
- (b) **'Link to show no change':** The price of the new item is linked into the index. The price change is assumed to equal the quality change and thus not included in the index calculation.
- (c) **Overlapping prices:** With overlapping prices the new item can be linked into the index. This assumes that the price difference reflects the value of any quality difference between the two items.
- (d) **Imputation:** The price development of the new item is imputed by the price development of similar items or groups of items.
- (e) **Monthly re-sampling and chaining:** The sample is updated regularly on monthly basis and a chained index is calculated on the basis of the matched monthly prices.
- (f) **Adjustment for quantity changes:** For items where the quantity change a proportionate change should be made in the price.
- (g) **Experts judgement:** Persons with detailed product knowledge are requested to value the difference between the new and old product, and appropriate adjustment are made in the recorded prices.

Hedonic regressions

74. Hedonic regression is a special way to adjust for quality changes in which the relationship between the price of an item and its price determining characteristics is estimated in a regression equation. The estimated coefficients reflect the influence on the price from the price determining characteristics. For a product with new values of the characteristics, for example a PC with more RAM, the quality adjusted price can then be determined. Hedonic regressions are relatively resource demanding, require relatively large amount of observations and regular updating.

75. Examples of methods a) – e) are provided below where it is assumed that the elementary index is calculated as a monthly chained index.

Direct comparison (include the whole of the price change)

	March	April	May
Item A	200	220	
Item B		220*	190
Monthly index	100,0	110,0	87,4
Chained index	100,0	110,0	96,1

* Adjusted price

76. In May item A is replaced by item B. It is assessed that there is no quality difference, and the whole price difference is included in the index by using the price of A as previous price of B.

Link to show no change'

	March	April	May
Item A	200	220	
Item B		400*	400
Monthly index	100,0	110,0	100,0
Chained index	100,0	110,0	110,0

* Adjusted price

77. In May item A is replaced by item B. It is assessed that the whole of the price change is due to a quality change, and B is included in the index with an unchanged price from April to May, so that the replacement does not affects the index.

Overlapping prices

	March	April	May
Item A	200	220	
Item B		400	420
Monthly index	100,0	110,0	105,0
Chained index	100,0	110,0	115,5

78. Overlapping prices are recorded for April. This allows the price of the new item to be linked into the index, assuming that the price ratios in April reflect only quality changes.

Imputation

	March	April	May	June
Item A	190	210	200	190
Item B	400	400	400	380
Item C	300	300		
Item D			240	240
Monthly index	100,0	103,4	97,6	96,6
Chained index	100,0	103,4	100,9	97,5

Note: The indices are calculated on the basis of geometric means

79. From March to April the index is calculated on the basis of the prices for items A, B and C. The monthly index for April-May is calculated on the basis of prices for A and B only. The new item D is included in the monthly index for May-June (calculated on the basis of A, B and D), which is then linked onto the value of the chained index in May.

Monthly re-sampling and chaining

	Month 0	Month 1	Month 2	Month 3	Month 4
Item A	a ₀	a ₁	a ₂	a ₃	
Item B	b ₀	b ₁	b ₂	b ₃	b ₄
Item C	c ₀	c ₁	c ₂	c ₃	c ₄
Item D		d ₁	d ₂	d ₃	d ₄

80. In this method the index is calculated by multiplying the monthly indices, each based on a set of matched prices including only the prices obtained in both periods. Using geometric means, the index in month 4 would be calculated as:

$$P_{0:4} = \left(\frac{a_1 \cdot b_1 \cdot c_1}{a_0 \cdot b_0 \cdot c_0} \right)^{1/3} \left(\frac{a_2 \cdot b_2 \cdot c_2 \cdot d_2}{a_1 \cdot b_1 \cdot c_1 \cdot d_1} \right)^{1/4} \left(\frac{a_3 \cdot b_3 \cdot c_3 \cdot d_3}{a_2 \cdot b_2 \cdot c_2 \cdot d_2} \right)^{1/4} \left(\frac{b_4 \cdot c_4 \cdot d_4}{b_3 \cdot c_3 \cdot d_3} \right)^{1/3}$$

81. The method assumes a competitive market and requires a relatively large number of observations.

IX. DATA EDITING

82. The data editing process comprises two main steps: detecting of possible errors and outliers, and validation and correction of data.

Errors and outliers 83. Errors may be falsely reported prices, or they may be caused by recording or coding mistakes. Possible errors and outliers are usually identified as observations that fall outside some pre-specified acceptance interval or are judged to be unrealistic by the analyst on some other ground.

'Inliers' 84. It may also be the case, however, that even if an observation is not identified as a potential error, it may actually show up to be false. Such observations are sometimes referred to as inliers. The reporting of unchanged prices by respondents is a particular source of error that potentially may cause a downward bias in the CPI if not corrected.

Validation 85. When a possible error has been identified, it needs to be verified whether it is in fact an error or not. This clarification can usually be made by asking the respondent to verify the price, or by comparison with the price change of comparable items. If it is in fact an error, it needs to be corrected. This can be done easily if the respondent can provide the correct price or,

where this is not possible, by imputation or omitting the price from the index calculation.

86. Sometimes exceptional price changes outside the acceptance interval are captured but verified as correct. Such outliers should be included, or otherwise omitted or imputed according to predefined practice.

87. The identification of extreme or unusual price changes can be accomplished by *non-statistical checking of input data*, *statistical checking of input data* and *output checking*.

Non-statistical checking of input data

88. Non-statistical checking can be undertaken by manually checking the input data, by inspection of the data presented in comparable tables, or by setting filters.

Statistical checking of input data

89. Statistical checking of input data compares, for some time period, each price change with the change for similar goods or services. Statistical checking or filtering may use different types of means or the median of the observed price changes on the basis of which an acceptance interval can be generated. Such intervals may also detect for example reported constant prices as potential errors if the general price trend is increasing.

Output checking

90. Checking by impact, or data output checking, is based on calculating the impact that an individual price change has on an index to which it contributes. This index can be an elementary aggregate index, the total index, or some other aggregate index. A maximum value for this impact can be set, so that only those price changes that cause an impact greater than this have to be reviewed.

X. DISSEMINATION

CPI is a key economic indicator and part of Official Statistics

91. The CPI is one of the most important economic statistical series. It must be published and otherwise disseminated according to the policies, codes of practices and standards for such data. Above all, the CPI should meet the *Fundamental Principles of Official Statistics* (UN 1994).

92. According to the *CPI Manual*, the CPI should therefore be:

- Released as soon as possible
- Made available to all users at the same time
- Released according to pre-announced dates
- Released separately from ministerial comments
- Made available in convenient form for users
- Accompanied by methodological explanation
- Backed up by professionals who can answer questions and provide further information

93. For the index to be widely accepted, it is essential that the decisions about how to compile the CPI are based on strictly professional grounds and that methods are documented and made available to the public.

* * * * *