

STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE

INTERNATIONAL LABOUR ORGANIZATION

CONFERENCE OF EUROPEAN STATISTICIANS

Joint ECE/ILO Meeting on consumer price indices
(Geneva, 24-27 November 1997)

Item 8 of the provisional agenda

**SAMPLING DESIGNS IN CONSTRUCTING CONSUMER PRICE INDICES:
CURRENT PRACTICES AT STATISTICAL OFFICES**

Paper submitted by Statistics Netherlands ¹

1. Introduction

1. As part of the statistical programme of the Statistical Office of the European Communities (Eurostat), Statistics Netherlands is carrying out research into several areas concerning the harmonised consumer price indices (CPIs). The main objective of this research is the production of national CPIs on a comparable basis. CPIs which differ on account of differences in the concepts, methods or practices used in their definition and compilation are not considered comparable.

2. One of the research topics is sampling. It is a well-known fact that major differences exist in the sampling methods and practices used in the construction of CPIs between the Member States of the European Union (EU). Differences in sampling practices can be regarded as a possible source of non-comparability of CPIs. Before we can perform a study into the systematic differences between CPI estimates based on different sample designs, it is necessary to collect information on the sampling practices in the fifteen EU countries. National statistical offices are asked to supply the data required. The questionnaire contained the following sections:

I. Sampling of geographical localities;

¹ Prepared by Mr. Martin Boon, Statistics Netherlands

- II. Sampling of outlets;
- III. Sampling of items ¹ at the central office;
- IV. Sampling of varieties ² in each outlet;
- V. Miscellaneous, viz. adjustment for non response, optimum sample allocation and references.

3. The main results of this survey, which took place in March and April 1997, are presented in this paper. Completed questionnaires were returned by twelve EU countries. Section 2 presents briefly, for each participating EU Member State separately, the sampling methods used in constructing CPIs. We also included relevant data about the United States, because the methods applied there are fully based on probability sampling and can be considered as the reference method. We have not considered comparison with other major countries such as Canada and Australia, because they apply similar non-probability sampling methods as the EU countries. In section 3 some general conclusions are drawn from the information received from the Member States. The questionnaire is given as an appendix.

2. Findings

4. In this section we present the sampling designs in constructing CPIs of the various EU countries (including the USA), listed in alphabetical order. It should be kept in mind that, due to the varying degree of the answers received, these descriptions differ in length.

2.1 Austria

Sampling of geographical localities

5. Geographical localities are selected by cut-off sampling. For each of the 9 Austrian 'Länder' the capital town is chosen. In addition, 11 towns with more than 20 000 inhabitants are included into the sample of localities.

Sampling of outlets

6. The outlets are selected by the local city administration together with the local statistical committee. Outlets for the centralised price collection are chosen by the central office. There is no special kind of list or register available for the outlet selection. The objective is to attain a sample of outlets which is representative for the market of the town in question. Street markets as well as not permanently open outlets are excluded from sample selection. The selection process can be characterised as judgmental. Prices are collected regionally from 4 200 outlets. When a selected outlet is permanently closed, it is replaced by another one. Further, new outlets are only introduced into the sample at base year revisions.

Sampling of items at the central office

7. In general the selection of the items is based on expert judgement. Only

when data on an auxiliary variable such as sales are available cut-off sampling is used. This concerns for instance cars, fruit, vegetables and public transport, which amounts to about 10% of the index. In total the CPI covers 710 items. Changes in the item sample are discussed in special committee meetings between the central office, the Chamber of Commerce and the Chamber of Labour. The item sample is updated every base year revision (i.e. five yearly).

Sampling of varieties in each outlet

8. The descriptions of the items are of a general nature, to enable price collectors to choose the type, quantity etc. in each outlet for regular pricing. The selection process of varieties can be described as a combination of judgmental selection, cut-off sampling and quota sampling. Every month about 80 000 price quotations are collected.

Adjustment for non response

9. Temporarily missing prices are imputed by extrapolating the observed price of the previous period using the change of the other prices for the same item.

Optimum sample allocation

10. The size of the outlet sample per town depends on the population size, while the number of chosen varieties per item depends on the actual CPI weight of the item in question.

2.2 Denmark

Sampling of geographical localities

11. For fresh food, clothing and footwear, which cover 17% of CPI weight, prices are collected by field representatives visiting the outlets in 30 cities. The total sample of the 30 cities is made up by the 5 biggest cities (which are selected with certainty) and 25 smaller cities (for which the selection is based on the judgement of experts). Prices for other goods and services are collected by questionnaires sent to outlets across the country.

Sampling of outlets

12. The retail sales register, the register of business units registered for VAT settlement and telephone books are used as sampling frame for the outlets. Mail order firms and street markets are not covered by the sampling frame. The selection of the 1 900 outlets is based on expert choice. The outlets in the 5 biggest cities are replaced regularly after 3-5 years, while the periodic replacement of the outlets in the other cities occurs every third year. Generally, the outlets reporting prices on questionnaires are replaced every third year.

Sampling of items at the central office

13. The choice of the 1 200 representative items rests on the judgement of the staff in the central office. The item samples are updated when existing items run out of stock or when new items become significant.

Sampling of varieties in each outlet

14. At each selected outlet a certain variety within the item specification is selected for pricing by the price collector. This selection is grounded on the judgement of the price collector or the shopkeeper. Each month approximately 25 000 price quotations are gathered for the all-items CPI.

Adjustment for non response

15. Missing prices are, as a general rule, omitted from the index calculation.

Optimum sample allocation

16. One of the criteria for determining the optimal size of the CPI sample is the cost of the price collection.

2.3 Finland

17. The price collection for the CPI is divided into four groups. For groceries and other items consumed daily (covers about 22% of CPI weight), field representatives gather price information from selected outlets in 107 municipalities. Secondly, prices of household appliances, clothes, various services and leisure-time goods (21% of CPI weight) are collected from 14 major cities (regionally centres). Thirdly, prices of various items are picked from catalogues of mail order firms and price lists of so-called discount stores. Finally, prices are collected centrally for uniformly priced services and other fairly easily obtainable goods (57% of CPI weight).

Sampling of geographical localities

18. The municipalities and major cities are selected using stratified sampling with probabilities proportional to population size (PPS) in each stratum (region).

Sampling of outlets

19. The sample of grocery stores in each selected municipality is drawn from the Finnish Business Register, supplemented with Nielsen data on grocery size. Outlet sampling was carried out by using the PPS method, with outlet turnover as size measure. Further, one has to check whether the selected outlet is in operation and the assortment of the selected outlet. The reason behind this check is the possibility of imperfections in the sampling frame of outlets. The outlets for household appliances etc. in the selected major cities are chosen judiciously by the price collectors. In other words, the outlet selection is based on the impression of price collectors on the popularity of outlets in their own city. In addition, mail order firms and discount stores are selected by cut-off sampling. Street markets, door-to-door selling, network marketing (like tupperware) are excluded from sample selection. The updating of the outlet sample is done at every CPI revision. Whenever an outlet has gone out-of-business, a

new outlet is chosen of the same type. Excluding centralised collection, prices are collected from nearly 2 000 outlets.

Sampling of items at the central office

20. On the basis of data from the Household Expenditure Survey the CPI experts of the central office choose representative items. The current CPI consists of 401 items. The price collectors are given the opportunity in each outlet to select the actual brand within the limits given in the item description. The package size and certain technical characteristics are given as a guide to price collectors. The item sample is changed according to the actual consumption at every CPI revision.

Sampling of varieties in each outlet

21. The price collector is free to select a certain variety in each outlet within the item definition which is provided by the central office. This selection process can be characterised as judgmental sampling. A total of about 43 600 prices are collected monthly in the entire country for the CPI.

Adjustment for non response

22. Temporarily missing prices are neglected in the CPI calculation.

Optimum sample allocation

23. At the moment Statistics Finland is carrying out a research project on sample allocation in the CPI. The proposed changes will be implemented this year. See Kinnunen (1993) for further details about the sampling designs in the CPI.

2.4 Germany

24. Although attempts are made to apply the same methods to the CPI for East-Germany as are used for the CPI for West-Germany, variations in the sampling practices could not be avoided. These differences are pointed out below, as they occur. On account of the federal structure of Germany the Federal Statistical Office (FSO) and the 16 Statistical Offices of the 'Länder' have different responsibilities in the execution of the CPI sampling. Germany uses non-probability sampling at all selection stages.

Sampling of geographical localities

25. Municipalities with less than 5 000 respectively 10 000 inhabitants in West-Germany and East-Germany are cut off or excluded from sample selection. For each 'Land' the municipalities are stratified according to population size. The sample sizes of the strata are fixed in advance in proportion to their share of total population. Then within the strata, municipalities (190 altogether) are chosen judiciously by the 'Länder' Offices.

Sampling of outlets

26. The chosen municipalities carry out the selection of the outlets in co-operation with the 'Länder' Offices according to FSO guidelines. Retailer dealing with luxury items and weekly markets are excluded from sampling selection. The number of outlets observed varies with the size class of each municipality. The location (inner city, suburb) and the type (independents supermarkets, etc.) are determining factors for the outlet selection based on the judgement of experts. Prices are collected from 22100 outlets (excluding tenants and landlords) for whole Germany. Outlets are replaced in connection with the regular updating of the item basket every five years. New outlets are introduced to replace closed ones.

Sampling of items at the central office

27. The selection of items can be characterised as judgmental sampling. The FSO selects and centrally specifies representative items from the Household Expenditure Survey supplemented with turnover data. The item specifications, especially durable goods, are broadly defined to facilitate the selection of best selling varieties at each outlet. The German CPI covers about 750 items. All items are reviewed and updated as necessary every five years. The proposals for the inclusion of new items are based on market information on changes in consumers' buying habits.

Sampling of varieties in each outlet

28. The price collector chooses in combined effort with the outlet the best selling variety within the range of each item specification, i.e. by cut-off sampling. In the German CPI about 400 000 price quotations are collected monthly. Once a variety has been chosen for a particular outlet it should be used for price quotations for as long as possible.

Adjustment for non response

29. In case of missing prices the previous price is extrapolated by the average change of the reported prices of the same item in the same municipality. It is not allowed to impute a missing price for longer than two consecutive months.

Optimum sample allocation

30. As in the German CPI only non-probability sampling is applied an exact definition of the criteria for sample allocation is not possible. In general, one tries to get a substantial sample size for non-homogeneous items, in which homogeneity is measured by the variance of the price index for an item.

2.5 Greece**Sampling of geographical localities**

31. Greece is distributed into 13 regions on the basis of the following criteria: geographic characteristics (mainland, island), character of the terrain (mountainous, flat) and level of economic growth. The towns with less than 10 000 inhabitants are excluded from sample selection. In each region 1-2 towns are

selected using expert judgement. In the whole country 17 towns were selected.

Sampling of outlets

32. For the selection of outlets regional business registers are used. The sampling frame does not include shops with low quality items, luxury shops and street markets which sell non-food items. The volume of sales and the geographical location are the basic criteria for the outlet selection based on expert judgement. An extended update of the outlet sample is made at each revision of the index (i.e. every 5-6 years). The number of outlets from which prices are collected (except rents) amounts to 3 200.

Sampling of items at the central office

33. The selection of the items is judgmental and takes into account the following criteria:

- the representativeness of the price movement for a given item group;
- the volume of sales of the item;
- the availability of the item on the market for a reasonable length of time.

An extended update of the item sample is made at each base year revision. The current CPI consists of 600 items.

Sampling of varieties in each outlet

34. The item specifications which are provided by the central office to the price collectors are not tightly described. Within each outlet, the price collector has to choose the particular variety to be priced. The choice of the variety is based on the same criteria as the item selection. The number of price quotations for all items (excluding rents) amounts to about 20 000.

Adjustment for non response

35. For the non-seasonal items the treatment of the missing prices depends on the length of the period in which the item is not sold in the outlet in question. In case of an absence period up to 1-2 months, the last regular price observation is carried forward, while for an absence period exceeding 1-2 months the missing price is imputed by extrapolation using the change of the other prices for the same item.

Optimum sample allocation

36. The size of the local market as well as the cost connected to the local price collection are determining factors for the optimal size of the outlet sample.

2.6 Ireland

Sampling of geographical localities

37. All towns with 10 000 or more inhabitants are selected with certainty. The set of remaining towns is subdivided into strata according to population size. In each stratum a simple random sample of towns is taken. In total 82 towns are selected for price collection purpose.

Sampling of outlets

38. There is no register which can be used as sampling frame of outlets. The outlet selection is based on the judgement of price collectors with the aid of the central office. The criterion for the selection of outlets is that the outlet should be representative of the locality. Mail order firms and street markets are excluded from outlet selection. Prices are collected regionally from 3 898 outlets. The outlet sample is in principle fixed in time between the five-yearly base year revisions. Only in special cases, such as the building of a new and large shopping centre and price collection problems with a particular locality, the outlet sample will be re-examined.

Sampling of items at the central office

39. The items are judgmentally sampled by the central office on the basis of their relative importance, representative nature and suitability for regular pricing. The selection is based on data derived from household expenditure surveys supplemented by information sought from supermarket head offices, retail and trade associations, magazines etc. The total number of items priced is 560. Up to the present the item sample has been fixed between base year revisions. The brands, size etc. may change but the basic item definition does not, unless the central office is made aware of a significant change.

Sampling of varieties in each outlet

40. The item specifications provided by the central office are of a general nature. The price collectors are free initially to choose a specific popular variety for regular pricing. The number of selected varieties is fixed in advance by the central office. This selection method can be called quota sampling. Approximately 45 000 prices are collected monthly from regional and central surveys.

Adjustment for non response

41. For prices collected regionally missing observations are neglected without re-weighting the other prices, while for centrally collected prices non-response adjustment takes place by means of re-weighting the observed prices.

Optimum sample allocation

42. There has been no recent work done on the optimisation of the sample sizes. The selection of towns is primarily historical. The size of the item samples reflects the importance of the attributed CPI expenditure weight. There

are no statistical determining factors for the optimal size of the outlet and variety samples.

2.7 Italy

Sampling of geographical localities

43. Prices are collected in 82 provincial capitals throughout the country.

Sampling of outlets

44. Within each provincial capital, the selection of outlets is carried by the local statistical office ('Uffici Comunali di Statistica') according to the guidelines of the central office. The outlet sample takes into account the various outlet types (department stores, supermarkets, traditional shops etc.) and the different area types (city centres, suburbs, outlying areas). The sampling frame is based on the outlet list of the chamber of commerce, census data and other local sources. Mail order firms and teleshopping firms are not included in the sampling frame. In practice, the outlets with the highest total sales or the highest sales of the item (group) in question are chosen (i.e. cut-off sampling). Selected outlets are replaced if they cease trading and/or if they change in type. At each base year revision the outlet sample is updated. At present about 25 000 outlets are used for local price collection.

Sampling of items at the central office

45. The choice of the so-called representative items is based on the judgement of experts. When a selected item ceases to be sold, the item is replaced by another. At each base year revision the item sample is revised. About 920 items are included into the current CPI.

Sampling of varieties in each outlet

46. In case of loosely specified items, the price collector chooses the most frequently bought variety in the outlet. About 300 000 price quotations are gathered monthly.

Adjustment for non response

47. If a price is missing for 1-2 months, the previous price observation is carried forward. If a price is unavailable for a longer period, then the outlet will be replaced.

Optimum sample allocation

48. The current sizes of the samples for the all-items CPI are not optimal. Research will be carried out on optimum sample allocation.

2.8 Luxembourg

Sampling of geographical localities

49. In general, retail prices are collected in 3 geographical regions: in the

north (centred on the cities of Diekirch and Ettelbruck), in the south (centred on the city of Esch-sur-Alzette) and in the centre of the country (centred on the city of Luxembourg). In addition, prices of municipal services are observed in a sample of 22 towns.

Sampling of outlets

50. The business register of Luxembourg is used as a sampling frame of outlets. Mail order firms, street markets and small non-specialist food shops are excluded from sample selection. The selection of the outlets is based on the judgement of experts. In principle the outlet sample is fixed in time, i.e. there is no periodic rotation. Whenever an outlet in the sample disappears, it is immediately replaced by a new outlet of the same type. Important new outlets are introduced into the sample. Prices are collected from 634 outlets.

Sampling of items at the central office

51. The selection of 6 744 items for which prices are collected is judgmental. The item specifications are detailed as regards variety, brand and model. They are the exclusive responsibility of the central office. Some 5 700 price quotations are collected monthly. The item sample is not regularly revised in a systematic way. If an item disappears from the market, it is replaced by another one with similar characteristics. Newly significant items are introduced into the index in December.

Adjustment for non response

52. Temporarily missing prices are treated according to either the method of carrying forward the previous price observation or the extrapolation method using the change of the other prices for the same item. In each specific case a decision is made which method is applied.

Optimum sample allocation

53. Currently, no statistical criteria are used to optimise the sample sizes.

2.9 Netherlands

Sampling of geographical localities

54. Municipalities that do not have at least one city or town with more than 10 000 inhabitants are excluded from the price survey. The remaining set of municipalities are subdivided into strata according to region and size class. In each stratum municipalities are selected with probabilities proportional to size (PPS sampling). The number of inhabitants for each municipality is used as size measure. The size of the resulting sample of municipalities is 100.

Sampling of outlets

55. There is not a single sampling design of outlets covering all items. One part of the items is surveyed locally by field representatives, while for the remaining items separate mail surveys are carried out centrally. Since the

sampling design varies from judgmental/cut-off sampling to PPS sampling in case of mail surveys, we consider only the field survey. The sample of outlets is to some extent based on probability techniques. Somewhere in the past a sample of outlets was drawn. The exact sampling design cannot be traced. Starting with the most recent base year (1990), the sample is post-stratified according to type of outlet (which is in fact an indicator of the type of goods sold) and size class (measured by number of employees). In principle the sample is fixed in time, so that we have a panel of outlets which remains unchanged. Whenever an outlet in the panel disappears (or refuses to co-operate any longer), a new outlet in the same stratum is sampled at random with equal probabilities from a business register. Mail order firms are excluded from sample selection. Further, the sampling frame has extensive overcoverage, so that the newly sampled outlet may turn out not to sell the item in question. In that case a new outlet is sampled. New types of outlets are only introduced into the index at base year revisions.

56. Each month, local prices are collected in about 10 000 outlets. Some 1 000 firms regularly receive mailed questionnaires (excluding rent survey). About 3 500 owners/administrators receive a rent survey questionnaire once a year.

Sampling of items at the central office

57. Generally the item selection is a two-stage procedure. In the first stage a number of item subgroups are chosen using the cut-off method. Only the subgroups with the largest market shares are selected. In the second stage one or more specific items are chosen from each subgroup by means of judicious sampling. The selection of these so-called representative items is based on the judgement of experts who have knowledge of the market in question. In order to measure pure price changes these items are described in detail, to ensure that exactly the same items are observed in all outlets. Prices are collected for about 1 200 items.

58. Just the same as with outlets, in principle a panel (of items) is used. This implies that the sample of items is not regularly updated between successive base year revisions. When a sampled item ceases to be sold, another one will be substituted. New items which are completely different from existing ones are introduced into the index between successive base year revisions.

Sampling of varieties in each outlet

59. For items where tight specification entails a considerable risk that an outlet does not sell the item exactly as specified or where price collectors cannot use them, loose specifications are used. In these cases price collectors usually choose the most frequently bought variety that fits the item description. Excluding rent, some 100 000 prices quotations are collected monthly.

Adjustment for non response

60. Prices which are temporarily not observable in certain outlets are imputed by the following method: extrapolation of the previous price by the change of

the other observed prices for the same item.

Optimum sample allocation

61. Criteria for determining the number of price quotations for different items are the expenditure weight and the price variation. More prices are collected from items with high price variation and items with large expenditure weight.

2.10 Spain

Sampling of geographical localities

62. Cut-off sampling is used for locality selection, i.e. the localities with the highest population size are included into the sample. Prices are collected in 130 localities for food, beverages and tobacco, while other item groups are surveyed in 70 localities.

Sampling of outlets

63. There does not exist a register which can be used as sampling frame of outlets. The selection of the outlets is based on the judgement of experts. Apart from forced replacements, the outlet sample is updated when price collector and/or inspectors detect new important outlets. Prices are collected in about 29 000 outlets.

Sampling of items at the central office

64. The selection process of items can be described as a combination of judgmental selection by experts and cut-off sampling taking into account the items with the highest sales. The item sample is updated when price collectors and/or inspectors receive signals from the market about the importance of new items. The current CPI covers 471 items.

Sampling of varieties in each outlet

65. The selection of the variety that fits the item specification is based on the judgement of the price collector. About 150 000 price quotations are gathered monthly.

Adjustment for non response

66. Temporarily missing prices are treated according to either the method of carrying forward the previous price observation or the extrapolation method using the change of the other prices for the same item.

Optimum sample allocation

67. The optimum outlet sample size is a function of the variance of price index, the CPI-weight and the cost for price collection. The optimum size of the variety sample depends on the specific characteristics of the item in question.

2.11 Sweden

68. The Swedish CPI is composed of several independent price surveys for different item groups. These surveys can be categorised as:

- daily necessities system, covering most food items (except fresh food such as fruit, vegetables, bread, and fish) and other daily items which are typically found in a supermarket (15% of CPI weight);
- local price system, covering footwear, furniture, household appliances, restaurant services and fresh food not included above (17% of CPI weight);
- clothing survey (4% of CPI weight);
- central price system, covering alcoholic beverages, transport, fuel, communication, housing, electricity, water, other services related to the dwelling (such as drainage system, refuse collection), recreation, health and medical care (64% of CPI weight).

Sampling of geographical localities

69. Only in a few central surveys geographical localities are selected in the first stage of the sampling. This concerns services related to the dwelling supplied by municipalities and transport by taxi. In the taxi survey there have been selected 19 municipalities.

Sampling of outlets

70. In Sweden outlet sampling is mainly done by probability methods. The Swedish Business Register is used as a sampling frame for the outlets in the daily necessities system, the local price system, the clothing survey and a few centralised price surveys. In general, a somewhat larger sample is selected than what is needed for the CPI, because it is possible that the assortment of some selected outlets does not contain the desired items. Purposely the following outlet types are left out of the scope of the CPI: food specialist shops (such as greengrocers, butchers, fish shops, bakeries), other specialist shops (such as furrier's shops, pet shops), mail order firms, street markets and repair workshops (for clocks, bicycles etc.).

71. The actual sampling procedure used in the price surveys (except the centralised surveys) is stratified PPS sampling. The number of employees is used as the size measure. The outlets are stratified by type of outlet (according to the Swedish Industrial Classification). In the central price system several sampling designs are applied in practice.

72. In the daily necessities system, the local price system, the clothing survey and the petrol survey the outlet sample is kept up to date. This is achieved by re-selecting (rotating) 20% of the sample every year. For several surveys of the central price system the outlet sample remains constant in time. Excluding the central price system, prices are collected from a sample of 72 outlets altogether.

Sampling of items at the central office

73. Item sampling is done by probability methods for daily necessities and furniture, whereas judgmental selection is used in several steps for clothing and item groups belonging to the local price system (excluding furniture). For the other item groups there is either no item sampling at all (all products are covered) or the sampling procedures used (e.g. transport and communications) are a mixture of probability-based and judgmental selection. For the daily necessities the actual procedure is PPS sampling of items with turnover as size measure, while in case of furniture simple random sampling is applied. In the CPI there are 318 items for which an index is calculated. New items are included when their CPI weights become significant.

Sampling of varieties in each outlet

74. The price collector is asked to find the best selling variety within the outlet in question, which satisfies the item specification provided by the central office. For the whole CPI (excluding the centrally collected prices) 22 800 price quotations are gathered monthly.

Adjustment for non response

75. If a variety is temporarily not available in a particular outlet, the missing price is neglected in the index calculation. However, if a variety is not available in the base period (December) and the price collector can find a variety that fits the item description in subsequent months, then the base price is imputed by using the change of the other observed prices for the same item.

Optimum sample allocation

76. An optimum allocation of the multistage sample is reached by using a numerical optimisation procedure based on the variances of price indices, costs for going to an outlet and costs for collecting the price of a variety in different types of outlets.

See Dalén and Ohlsson (1995) for further details about the sampling and variance estimation methods used in the CPI.

2.12 United Kingdom**Sampling of geographical localities**

77. Localities are defined as contiguous clusters of postal sectors, broadly representing a central shopping area and the areas where its shopping population lives. These localities are assigned to strata according to region, type of location (major centre, district centre and local centre) and shopping population size. The largest strata, that is major towns with shopping populations of over 500 000 and the five out-of-town shopping centres, are sampled with certainty. The numbers of localities to be selected in the remaining strata are allocated in proportion to total shopping population. Within each stratum, localities are selected using simple random sampling. The localities with a shopping population below 10 000 were excluded from sample selection. The number of localities where prices are collected is 146.

Sampling of outlets

78. For outlet sampling, a sampling frame is established through enumeration of the main shopping areas within each locality. Outlet enumeration yields a list of all outlets in the area specified together with the item groups which belong to their assortments. In addition, market stalls selling fruit, vegetables and fish are enumerated and included in the sampling frame along with retail outlets. Market stalls selling other goods and outlets which require a membership fee are not included in the frame. An outlet coding scheme is devised to link outlets with items.

79. Outlets of each item group are split by shop type (multiple, independent, or specialist) to form different strata. For each stratum, the required number of outlets is drawn either by simple random sampling or by sampling probability proportional to size. Roughly one fifth of the CPI weight is from items priced in outlets selected by PPS. The retail floor space is used as a size measure of the outlet. Any given outlet may be eligible for selection more than once if it sells items from more than one item group.

80. As outlets close, new ones are selected from the enumeration list. Since outlet enumeration is very costly, it is carried out only at five year intervals. The lists are updated if price collectors or auditors notice any major changes in the outlets.

81. At present around 18 000 outlets are used for local price collection. This does not include some of the major retailers who have central price policies, such as mail order firms. Prices from these firms are collected by the central office.

Sampling of items at the central office

82. Items are judiciously sampled at the central office. The selection is based on expert judgement that is backed up by expenditure data and market research data. The total number of items currently collected is 646. Each year the selection of all items is examined and compared to the latest expenditure and market research data. Old items are deleted and new ones brought in every January.

Sampling of varieties in each outlet

83. In general items are defined quite loosely so it is possible to select a number of different varieties to represent each item. Price collectors are told to choose popular varieties. Each month around 120 000 prices are collected. This includes both central collection by central office and local collection of prices throughout the country.

Adjustment for non response

84. When an item is either temporarily or permanently unavailable, the price data are excluded from the index calculations but no re-weighting is carried out.

The only data used in the index is that for which there is both a base price and a current price.

Optimum sample allocation

85. Some initial work on optimum sample allocation is carried out last year and is based on the sampling variances for the item price indices. More prices are collected from items with high price variability and fewer prices from items with lower price variability.

See Haworth (1996) and Skinner et al. (1997) for more details about the selection processes and the recent sample optimisation exercise.

2.13 United States

86. The US CPI sample design involves a four-stage full-probability selection process.

Sampling of geographical localities

87. For the CPI 88 geographical primary sampling units (PSUs) are sampled throughout the country in which prices are collected. One part of the PSUs is selected with certainty, while the remainder are selected with probability proportional to population size. Each PSU is a county or group of counties and can range in population from several million in large urban areas to about 2 000 in smaller urban areas. Rural areas are excluded from the CPI.

Sampling of outlets

88. The selection of the outlets is based upon a so-called point-of-purchase survey which is conducted in each selected PSU. This survey is a continuing household survey that obtains the outlets at which the items were purchased and the amount spent at each outlet for each item. Then, a sample of outlets is drawn with probabilities proportional to expenditures. Prices are collected from some 21 000 outlets. To enable the CPI to reflect current buying patterns, each year the outlet sample is rotated in 20 percent of the sampled PSUs.

Sampling of items at the central office

89. All consumption expenditures from the Household Expenditure Survey are classified into exclusive and exhaustive item groups. For each sampled PSU, items are selected from each item group with probabilities proportional to estimated annual household expenditures in the PSU in question. The current US CPI covers 370 items. The sample of items is rotated in about 20% of the sampled PSUs each year.

Sampling of varieties in each outlet

90. Within a sampled outlet, one specific variety for each item is selected using probability proportional to item's sales as reported by the outlet's respondent. The selection of a variety at each outlet can involve sampling in successive stages. For the all-items CPI, 80 000 price quotations are gathered

each month.

Adjustment for non response

91. To adjust for non response the observed prices are re-weighted to get population estimates for the item price changes.

Optimum sample allocation

92. The number of prices to be collected for each item are allocated so as to produce the smallest possible sampling error for the US all-items CPI within the overall available budget. This allocation requires a non-linear programming approach that considers the relative expenditure importance of each item, the unit variance, the components of the variance, the unit cost of data collection and processing. Items with greater expenditure weights require larger sample sizes than items with smaller weights. Similarly, items with large unit variances also require larger sample sizes to reduce their contribution to the overall CPI variance.

See US BLS (1993) for more details about the sampling designs in the CPI.

3. Conclusions

93. Tables 1, 2 and 3 summarise the way how the eleven participating EU countries and the USA draw the different samples used in the CPI. We can draw the following conclusions from these tables.

Sampling of geographical localities and outlets

94. Prices are collected in a sample of outlets. Sampling methods vary considerably from country to country. They largely depend on the diversity of prices, on the pattern of retail trade and on the commercial practices which are peculiar to each country. In the most EU countries a two-stage sampling design is applied. First a sample of municipalities is drawn and then in each selected municipality a sample of outlets is chosen. The volume of commercial activity, whether roughly estimated or based on observed values, is a factor which often comes into play. Thus, probabilities proportional to the number of inhabitants are frequently used in the first sampling stage. In addition, outlets are chosen according to their turnover, on either a probability or judgmental basis. Only four EU countries use probability techniques for outlet selection.

Sampling of items and varieties

95. With the exception of Sweden, in the EU countries the items are chosen at the central office using criteria based on representativeness, rather than probability sampling techniques. Generally the item selection is a two-stage procedure. In the first stage a number of item subgroups are chosen using the cut-off method. Only the subgroups with the largest market shares are selected. In the second stage specific items are chosen from each subgroup by means of judgmental selection. In some cases loose item definitions are specified by the central office. Then, within the outlets price collectors usually choose the most frequently bought variety that fits the item description.

Outlet coverage

96. Outlet types are covered to different extent in different member states. Most EU countries exclude mail order firms and non-food market stalls from outlet sample selection.

Optional replacement of items and outlets

97. Except for forced replacements in case of permanently missing prices, the item and outlet samples are usually revised when an index is re-weighted. In some countries even between re-weightings, the samples are revised either routinely at regular intervals, or when the importance in the market of new items or outlets becomes apparent.

Adjustment for non response

98. One part of the EU countries excludes the data of temporarily unavailable items from the index calculation (without re-weighting for non-response). In that case, the only data used in the index is that for which there is both a base (or previous) price and a current price. Another part of the participating countries impute missing prices by extrapolating the previous price.

Optimum sample allocation

99. The only EU countries that appear to make serious efforts at full probability sampling is the United Kingdom and Sweden. The other EU Member States use mainly non-probability techniques such as judgmental or cut-off sampling. The main problem with non-probability sampling is that it does not permit the precise statistical estimation of sampling errors of the CPI. The reason is that there is no way of knowing whether the dispersion in the sample data represents the dispersion in the population. Knowing the sampling error of the CPI is not only important for the users of the CPI but also for the CPI statisticians. Optimising the sample sizes at the different sample stages (localities, outlets, items, varieties) is only possible in the case of a probability sample. Countries who do optimise their sample allocation use generally the following criteria: variance of the price indices, cost of the price collection, travel cost and expenditure weights.

100. The US CPI sample design involves a multistage full-probability selection process. This design can be considered as the ultimate goal for each statistical office. The most difficult problem in selecting full probability samples for a CPI is the lack of appropriate sampling frames of items.

Endnotes

1. The term 'item' is used to mean any good or service which is provided by the central office to the price collector.
2. A variety is a more detailed description of the particular product selected in an outlet by the price collector within the item specification provided. d

Table 1. Sampling of geographical localities and outlets in EU countries and USA

Country	Localities		Outlets				
	Design	Size	Frame	Undercoverage	Design	Updating ^a	Size
Austria	cut-off	20	no	street markets temporarily open outlets	judgmental	every base year revision	4 200
Denmark	judgmental	30	yes	mail order firms,	judgmental	every third year	1 900
Finland	stratified PPS	107	yes	street markets street markets, door-to-door selling,	stratified PPS, judgmental,	every base year revision	2 000
Germany	cut-off, quota	190	no	network marketing shops for luxury items, weekly markets	cut-off judgmental	every base year revision	22 100
Greece	judgmental	17	yes	shops for luxury items, non-food street markets	judgmental	every base year revision	3 200
Ireland	stratified SI	82	no	mail order firms, street markets	judgmental, quota	every base year revision	3 898
Italy	unknown	82	yes	mail order firms, teleshopping firms	cut-off	every base year revision	25 000
Luxembourg	judgmental	22	yes	mail order firms, street markets, small non-specialist food shops	judgmental	in reaction to market signals	634
Netherlands	stratified PPS	100	yes	mail order firms	stratified SI, judgmental,	every base year revision	11 000
Spain	cut-off	130	no	unknown	cut-off judgmental	in reaction to market signals	29 000
Sweden	unknown	19	yes	specialist shops, mail order firms, repair workshops	stratified PPS	every year 20% by rotation	726
UK	stratified SI	146	yes	non-food market stalls, shops requiring a mem- bership fee	stratified PPS, stratified SI	five yearly	18 000
USA	stratified PPS	85	yes	unknown	stratified PPS	every year 20% by rotation	21 000

^a Except for (forced) replacement by comparable outlets of existing outlets which cease trading or refuse to co-operate any longer.

Table 2. Sampling of items and varieties in EU countries and USA

Country	Items		Size	Varieties		Number of price quotations ^b
	Design	Updating ^a		Design		
Austria	judgmental,	every base year revision	710	judgemental,		80 000
Denmark	cut-off judgmental	in reaction to market signals	1 200	cut-off, quota judgmental		25 000
Finland	judgmental	every base year revision	401	judgmental		43 600
Germany	judgmental	every base year revision	750	cut-off		400 000
Greece	judgmental	every base year revision	600	judgmental		20 000
Ireland	judgmental	every base year revision	560	quota		45 000
Italy	judgmental	every base year revision	920	cut-off		300 000
Luxembourg	judgmental	in reaction to market signals	6 744			5 700
Netherlands	cut-off,	in reaction to market signals	1 200	cut-off		100 000
Spain	judgmental cut-off,	in reaction to market signals	471	judgmental		150 000
Sweden	judgmental stratified PPS, SI,	in reaction to market signals	318	cut-off		22 800
UK	judgmental	yearly	646	cut-off		120 000
USA	judgmental PPS	every year 20% by rotation	370	multistage PPS		80 000

^a Except for (forced) replacement by comparable items of existing items which cease to be sold.

^b Excluding price quotations used in measuring housing items.

Table 3. Miscellaneous sampling issues in EU countries and USA

Country	Treatment of missing prices	Sample allocation criteria
Austria	extrapolation	CPI weight, population size
Denmark	neglect	cost of the price collection
Finland	neglect	planned
Germany	extrapolation	variance of the price indices
Greece	carrying forward, extrapolation	cost of the price collection
Ireland	neglect, re-weighting	CPI-weight
Italy	carrying forward	planned
Luxembourg	carrying forward, extrapolation	
Netherlands	extrapolation	CPI weight, price variation
Spain	carrying forward, extrapolation	population size, variance of price indices due to the sampling of outlets CPI-weight, costs for price collection
Sweden	neglect	variances of price indices due to the sampling of outlets and items, travel costs, costs for price collection
UK	neglect	variance of price index due to the sampling of outlets
USA	re-weighting	variances of price indices due to the sampling of localities, outlets and items, weight, costs of data collection and processing

4. References

- Dalén, J. and E. Ohlsson, 1995, Variance estimation in the Swedish consumer price index. *Journal of Business and Economic Statistics* 13 (3), pp. 347-356.
- Haworth, M.F., 1996, Re-engineering data production and measuring quality in the UK retail prices index. *Proceedings of the Annual Research Conference* (US Bureau of the Census, Washington DC).
- Kinnunen, A., 1993, Consumer price index 1990=100, methodology and practice (Statistics Finland, Helsinki), pp. 44-47.
- Skinner, C.J., C.J. Holmes, S. Purdon, M.F. Haworth and R. Beaven, 1997, Allocation of price data collection in the retail prices index. Unpublished paper (Office for National Statistics, London).
- US BLS, 1992, BLS Handbook of Methods, Bulletin 2285 (US Government Printing Office, Washington).

5. Appendix: survey on the sampling design in constructing consumer price indices

A. Sampling design for the all-items CPI

Sampling of geographical localities

1. With which of the sampling designs mentioned in the appendix can the selection method of geographical localities be characterised? More than one sampling design may apply (please specify the total CPI weight for which every sampling design is applied).
2. Please give the number of geographical localities in which prices are collected.

Sampling of outlets

3. What kind of list or register is used as sampling frame? Please indicate whether a point-of-purchase survey is carried out for selecting where to collect prices.
4. Indicate which outlets belonging to the target population are not presented in the sampling frame because of undercoverage (for instance mail order firms or street markets).
5. With which of the sampling designs mentioned in the appendix can the selection process of outlets be characterised? More than one sampling design may apply (please specify the total CPI weight for which every sampling design is applied).
6. Describe how is dealt with changes in the outlet samples. Indicate whether the outlet samples are regularly updated (for instance by periodic rotation).
7. Please give the number of outlets from which prices are collected.

Sampling of items at the central office

8. With which of the sampling designs mentioned in the appendix can the selection process of items be characterised? More than one sampling design may apply (please specify the total CPI weight for which every sampling design is applied).
9. Describe how is dealt with changes in the item samples. Indicate whether the item samples are regularly updated.
10. Please give the number of items for which prices are obtained.

Sampling of varieties in each outlet (in case of loose item specifications provided by the central office)

11. With which of the sampling designs mentioned in the appendix can the selection process of varieties in outlets be characterised? More than one sampling design may apply (please specify the total CPI weight for which every sampling design is applied).
12. Please give an indication of the number of price quotations which are gathered monthly.

B. Miscellaneous

Missing prices

13. Please describe the treatment of missing prices (that means the adjustment for non-response) for the all-items CPI.
 - A. neglect (or omission) of the missing prices (without re-weighting);
 - B. imputation (replacement with a suitable estimate) of the missing prices:
 1. carrying forward the previous observation;
 2. extrapolation by using the change of the other prices for the same item;
 - C. re-weighting (or raising) of the non-missing prices (for instance with post-stratification);
 - D. other (please specify).

Optimum sample allocation

14. Indicate the criteria for determining the optimal sizes of the locality, outlet, item and variety samples for the all-items CPI.

References

15. Please mention specific references (internal notes or special sections in publications) on the description of the sampling and variance estimation methods used in the CPI and if possible enclose copies of these references.

Appendix: list of sampling designs

- A. simple random (or equal probability) sampling (SI);
- B. stratified sampling with SI sampling in each stratum (please specify the stratification variables);

- C. sampling with probability proportional to size (PPS, please specify the size measure);
- D. stratified sampling with PPS sampling in each stratum (please specify the stratification variables);
- E. non-probability (or representative) sampling:
 - 1. judgmental (or judicious) sampling (the selection of elements is based on the judgement of experts);
 - 2. cut-off sampling (the elements with the highest sales or value of other auxiliary variable are included into the sample);
 - 3. quota sampling (a priori fixation of the number of elements; the selection of the elements is delegated to price collectors);
- F. other (please specify).