PRICE COLLECTION AND VALIDATION

<div style="text-align: right; font-size: 2em;">**5**</div>

Introduction

5.1 Several factors will determine the choice of which price collection methods a national statistical office (NSO) uses, taking into consideration efficiency, accuracy, and representativity of consumers' purchasing patterns. For example, local price collection is costly but can have the advantage of covering a wide range of locations and items, particularly for food, alcohol, tobacco, and durable goods such as clothing, furniture, and electrical products. On the other hand, central collection, either at the head office or in regional offices, can be cheaper and may be used for products where there are national pricing policies, as for rail fares, or where prices cannot be observed directly in retail outlets, such as for many professional services. With regard to representing consumers' purchasing patterns, the price collection method also needs to reflect methods of shopping. For instance, internet purchases or goods purchased through mail order and catalogs need to be properly reflected in the sample.

5.2 Price collection will therefore require different practical solutions in different countries according to local circumstances. Thus, the most appropriate sampling and survey methods, and the best data sources for price collection will depend on the structure of retailing with regard to the characteristics of outlets, their geographical spread, and the range of goods and services available to purchase, plus the shopping behavior of the households covered by the index. The compiler should always be guided by the principles and objectives of the price index being compiled, as addressed in earlier chapters.

5.3 Price collection is becoming increasingly multimodal with prices being web scraped from the internet or obtained from scanner data, as well as being collected by hand from outlets and by telephone inquiry. Issues of coherence can arise when integrating price data from different sources, with relevant consequences. The internet prices charged by online retailers can differ from the prices charged by traditional outlets. Similarly, the internet prices obtained from a retail chain's website and the associated sales volume can differ from the corresponding in-store prices. The prices collected from each source should be weighted according to the respective values of sales via the different channels to ensure the index properly represents the purchases of the target population. Moreover, products representative of in-store sales may not be representative of internet sales and vice versa. Constructing separate elementary aggregates to represent different outlets or outlet types, and aggregating the elementary aggregates using explicit weights relating to the respective sales can help to ensure a balanced sample. For example, prices from a sample of bakeries (using price collectors), prices for bread from supermarkets (using

scanner data), and prices obtained via web scraping have to be aggregated in a correct and coherent way taking into account the weights of the different channels or sources. Ideally, indices per data source should be weighted and prices from different sources should not be mixed; however, in practice, this will depend on the availability of data to develop weights.

5.4 The activity of price collection needs to be carefully planned and a number of organizational options can be followed. The focus of this chapter is on traditional forms of price collection where price surveys are undertaken in outlets, using paper forms, handheld computers, or tablets, and supplemented by prices obtained centrally from postal, email, or telephone inquiry. However, the general principles described in paragraphs 5.15–5.49 apply to all types of price collection.

5.5 A separate section details the collection of prices on the internet (paragraphs 5.194–5.208). The use of scanner data is discussed in Chapter 10, considering the conceptual and practical issues associated with the use of these data.

5.6 The relative advantages of central and local price collection will depend on individual country circumstances. A further option involves the outsourcing of certain elements of price collection. In reviewing these options, this chapter covers the relationship between the field and the head office, and the handling of the flow of information.

Organization Options

Local Price Collection

5.7 Local price collection involves collectors visiting individual outlets to collect prices for several goods and services. This is the predominant method of price collection in most countries, although more countries are now starting to use other methods. The range and number of outlets visited and the types of goods and services priced vary among countries.

5.8 Although the precise method of local price collection varies, each price collector is usually responsible for collecting prices from a certain location or from certain types of outlets. Collectors will visit the same outlets in each collection period to attempt to price the same items. Through this type of arrangement, price collectors can build up effective relationships with retailers and gain specialist knowledge.

5.9 There are several important criteria relating to the conduct of the collection:

(1) Collectors should be appropriately dressed and be polite—they are representing the NSO.

(2) Collectors should carry identification to confirm their role and status.

(3) Collectors should make themselves known to the retailer or store manager when they arrive, and before they begin collecting prices.

(4) Collectors should comply with any request from the outlet staff if it does not invalidate the sample. For instance, if the store is very busy, the collector may comply with a request to return later in the day.

(5) The collection should be carried out as quickly as possible, causing minimal disruption to store business.

Central Price Collection

5.10 Centrally collected prices are obtained from the central offices of major retailing chains with national pricing strategies. Branches of these chains may be excluded from the local collection if data can be collected more effectively centrally. Data suppliers may provide information on paper forms, or by sending spreadsheets to the NSO by email, fax, or on external media such as a flash memory stick. Mail order catalogs can also be treated in the same way. These prices are then combined with those for the same items from the local collection. Prices can also be collected centrally online or by using web scraping or scanner data (see paragraphs 5.194–5.208 and Annex 5.6 and, for scanner data, Chapter 10).

5.11 Price data for services or fees may be collected centrally from organizations such as trade associations or national or local governmental departments. This is often the case for tariffs, for example, for gas and electricity, or for telephone and internet charges, where data from administrative sources on pricing structures and purchasing patterns are required (see Chapter 11).[1] Whenever possible, centrally collected prices are obtained from one central source, although contact with regional or competing companies is needed if there are local variations. Data may be requested in writing, by telephone, or may be delivered automatically because the NSO is on a provider's mailing list. Providers may send a full price list or tariff sheet, from which the relevant prices will be extracted by the consumer price index (CPI) staff, or just the prices of the items specified in the data request. It is good practice to confirm price quotations by some form of written documentation. When collecting prices from the internet, screen prints can be useful as displayed prices can frequently change (see paragraphs 5.194–5.208 relating to collecting prices online). Frequency of inquiry can vary across the range of items depending on when prices are known or expected to change. The most common frequencies are monthly, quarterly, and annual, but there are also instances of collecting as frequently as weekly or even daily. The frequency depends on price volatility. For regulated and fixed prices, collection could occur on the date when the change becomes effective. For instance, this may be the case where tariffs for gas, electricity, and water change once

a year on a set date. Internal checking procedures need to be in place for centrally collected prices, because these prices can often represent items with a relatively high expenditure weight. Credibility checking should be carried out to test the reliability of the collected prices, for instance, by comparing whether the price change relative to the previous price looks reasonable compared with historic evidence while taking into account sales prices.

The Principles of Price Collection

5.12 The procedures for price collection should follow the general concepts underlying the CPI and should reflect the expenditure of the target population.

Defining the Price

5.13 Given that the aim of the CPI is to measure price inflation, prices are defined according to what consumers pay in the marketplace (that is, the actual transaction prices of goods and services purchased by the consumer, including all taxes). However, in practice, the advertised price is usually selected because it can be readily observed. Practical difficulties stem from the existence of discounts offered when stock is sold off or when end-of-line, damaged, or special stock is brought in at sale time. In these cases, special procedures apply. Another complication that can arise is bargaining, where a price might not be displayed or where goods and services are purchased on the black market. Procedures to address these situations are discussed in paragraphs 5.24–5.29.

Nontransaction Prices

5.14 There are some exceptions that deviate from the stated aim of measuring actual transaction prices. The most notable example is the treatment of owner-occupied housing services, where alternative conceptual approaches require different measurement methodologies, and the choice of conceptual treatment depends on the main purpose of the index. For some services, the transaction price may be represented by a tariff or daily or hourly rate. More details on owner-occupied housing and tariffs are discussed in Chapter 11.

Catalog and List Prices (Not Mail Order)[2]

5.15 Catalog or list prices provided by the supplier to the retail outlet are in many cases identical to the transaction price. However, the catalog or list price may only be the recommended price and not the actual price the item is sold at. Even where it is supposed to be the actual price charged, retailers may not always comply with the prescribed prices. Although the use of catalog prices is contrary to the principle of recording transaction prices, in practice it can be a cost-effective method of obtaining prices. Therefore, catalog or list prices can be used, but their reliability should be confirmed periodically.

Price Reductions and Related Issues

5.16 Transaction prices may differ from advertised prices, for example, if at time of purchase a discount is

---

[1] There are a number of common issues associated with the use of administrative sources for the use of listed prices in the CPI, including coverage, definitions, data quality, and resilience of supply. For instance, does the source relate to the definition of the population being measured (transaction prices and sales to private households)? What is the quality of the data source (up to date with no errors)? How resilient is the supply (appropriate gateways available to access the data in an appropriate and reliable form on a regular and timely basis)?

[2] The pricing of goods and services ordered by mail can be conducted by reference to mail order catalogs. The index will need to include posting and packaging costs.

offered. In practice, however, discriminatory discounts, which are available only to a restricted group of households (as opposed to nondiscriminatory discounts that are available to all), are generally excluded on principle. For example, money-off coupons and loyalty rewards for previous expenditure are normally ignored and the nondiscounted price is recorded. Also, it may be difficult to obtain the price paid if it is subject to individual bargaining. The following practices are recommended to address different types of price reductions. Price collectors should make extensive notes on the situations confronted so that decisions on how to treat specific cases can be reviewed and confirmed at the head office. In general, if on the day scheduled for price collection, the price of an item is reduced due to sale or promotion, the reduced price is collected even if the sale or promotion is only for one or two days.

(1) *Discounted prices*: should only be included if generally available to everyone with no conditions attached; otherwise, the nondiscounted or unsubsidized price should be recorded. The general practice is to ignore money-off coupons and loyalty rewards. A judgment needs to be made, however, relating to the interpretation of "generally available." For instance, reduced prices for payment by direct debit may be collected depending on the extent to which consumers have access to, and use, of this service. A judgment is required on the threshold to be set for access, above which the discounted price is included in the index. Alternatively, for tariffs, different payment methods may all be priced individually (for example, separate data collection for payments for electricity by cash, direct debit, and prepayment) and the prices weighted and aggregated to form a single price index for that item.

(2) *Price discrimination*: discounts systematically available only to a restricted group of households should be disregarded because they are discriminatory, unless they are significant and are available either to the vast majority of the population or to identifiable subgroups who qualify for such discounts on the basis of demographic or other characteristics not requiring action by the individuals concerned at the time of purchase. If included, they should be treated as stratification or coverage issues in item sampling. A judgment is required from the compiler. Examples of price discrimination generally included in the CPI may include lower prices offered to senior citizens (for example, discounted travel or haircuts) and discounts for people who receive state benefits. An example where prices are not universally available and judgment is required is when a nominal or token membership fee is required by the retail outlet. In these cases, the take-up of such membership, which is widely available, needs to be considered with regard to thresholds and general spending patterns of the consumers and the conditions placed on membership that may make it restrictive (for example, minimum levels of purchase). See paragraphs 5.24–5.29 for a discussion on price bargaining.

(3) *Sale or special offer prices*: should be recorded if they are temporary reductions on items likely to be available again at normal prices or are stock-clearing sales (for example, January sales or summer sales). How-ever, before designating a price as a "sale" price, special care should be taken to ascertain that there is a genuine sale with price reductions on normal stock. On occasion, stock is continually sold below the recommended retail price or advertised as a special offer even though these prices are available all year. In such cases, prices should not be considered as sale prices, but should still be collected. Special purchases of end-of-season, damaged, or defective goods should not normally be priced, as they are likely not of the same quality as, or comparable with, goods previously priced and are unlikely to be available in the future. If the special offer is limited to the first customers, the item should not be priced, as the offer is not available to everyone. Introductory special offers may be included if they are available to all. However, given the need to price the same "basket" each month, such offers will not be chosen as representative items, unless they have sufficient volume of sales to be considered representative and are introduced at the time of an update of the basket, or when a replacement item needs to be chosen. Discounts on goods close to expiry date should be disregarded or treated as specification or quality changes.

(4) *Bonus offers, extras, and free gifts*: prices for items temporarily bearing extra quantities (for example, 30 percent extra free) should be adjusted to take account of the increased quantity unless it can be determined that the extra quantities involved will not be wanted by most consumers, will not have influenced the decision to purchase, and will not be consumed. Similarly, "2 for the price of 1" offers should be included. The reasoning behind this is that, if the CPI is tracking the price of a specific product, say a 330-milliliter can of a diet soda, and the offer is two cans for the price of one, the consumer will always take the two cans representing a 50 percent discount on a single can. However, an offer of three cans for the price of two would be disregarded because the offer requires the purchase of two cans and the item specification is for a single can. Money-off coupons for future purchases are usually disregarded as these may not be used or wanted. Free items with other purchases (for example, a free gift with every product purchased) are generally disregarded. For example, free gifts such as plastic toys in cereal boxes should be ignored because they are not included in the list for price observations; it is the price to be paid for the cereal in the box that is relevant. Also, such free gifts can be difficult to value. Similarly, receiving a free toothbrush when buying a bottle of mouthwash should be disregarded because it would be difficult to value the free gift and the consumer may only want the mouthwash and not the toothbrush. The treatment of bonus offers, extras, and free gifts is subject to interpretation on a case-by-case basis. Collectors should be aware that temporary "special offer" weight changes (X percent extra free) could become a permanent weight change (for example, cans of alcoholic drinks changing size from 440 milliliters to 500 milliliters) and should feed the information back to the head office as they become aware of it. In this way,

79

CPI staff can issue new or amended guidance to price collectors about item specifications.

(5) *Stamps*: sometimes purchasers are given special stamps, which can be accumulated and subsequently exchanged for goods and services. If a discount is available as an alternative to such stamps, then the discounted price should be recorded. Otherwise, the stamps should be disregarded.

(6) *Trade-ins*: in general, the price reduction obtained by trading in an old item (for example, a car) compared with the nominal full price should be ignored. This treatment follows convention, as the transaction essentially relates to a second-hand good and only the service charge levied by the outlet in buying and selling the good comes under the scope of the index. In practice, however, the situation is not so clear-cut. For instance, a garage may give a discount that is greater than the retail value of the traded-in car and, therefore, in effect gives a genuine discount on the new car. In many cases, discounts from trade-ins are difficult to evaluate. The trade-in value may be negotiable in each case, and the full nominal price—which is used as the benchmark against which the discount is measured— may not be known. It is therefore generally best to report the list price or asking price.

(7) *Sales taxes*: when a tax is not included in the price of individual items in an outlet, but is added when the customer pays for the item, care must be taken to record the price including tax. To make sure of this, with items for which the price is normally quoted pre-tax, and in areas where a general sales tax is added to the bill, the price collection forms should require the collector to indicate whether or not the price recorded includes the tax, as a price check, so that it can be added where necessary.

(8) *Tips for services*: if a compulsory service charge is included, for example, on a restaurant bill, only the compulsory amount should be included in the price, but not any additional discretionary tips. For services that are free in principle, but that in practice can rarely be obtained without a tip, or where tipping at a standard rate is the common practice, such tips may be added to the specified price. However, there is no agreed-upon convention.

(9) *Regular rebates or refunds*: should only be considered when attributable to the purchase of an individual identifiable product and granted within a time period from the actual purchase, and are expected to have a significant influence on the quantity buyers purchase. For example, money-back deposits on bottles should be deducted from the price if these are a sufficient incentive for returning the bottle, whereas money-back offers on lawn mowers after a five-year period should be disregarded. In all cases, a consistent decision for each item must be applied over time. Decisions about the treatment of rebates are made on an individual basis. They may reflect income rather than expenditure changes and may require different treatment from that used by national accounts.

(10) *Irregular rebates or refunds*: as with regular rebates or refunds, these should only be considered when they apply to the purchase of an individual product and are granted within a time period, and are expected to have a significant influence on the quantities purchasers are willing to buy. Loyalty rebates or coupons associated with previous expenditure at the outlet, to be used for similar or other purchases, should be disregarded as they are conceptually out of scope because they are discriminatory relying on previous purchases. One-off rebates should be disregarded as they do not relate to the specific time period of the consumption and are unlikely to affect levels of consumption. They are viewed more as a source of additional income.

(11) *Credit card and other payment arrangements involving interest, service charges, or extra charges*: charges incurred as a consequence of failing to pay within a specified period of time from the purchase should be disregarded. For example, zero or positive interest loans granted to finance a purchase should be disregarded when determining the price. These charges come under financial services.

(12) *Cash back*: whether to include a cash back offer depends on the circumstances. Some large retailers offer credit card accounts to their customers. In this case, a cash back offer could be considered a form of discount. Each time consumers use the retailer's credit card, they could earn a percentage of the total amount paid in the form of cash back. For example, if the card pays 2 percent cash back and the customer spends £100 in the retail outlet, they will earn £2. Most cash back cards credit the amount earned by the customer onto their statement, thereby reducing their credit card bill. Like loyalty cards, data collectors should collect additional information to determine what proportion of consumers use the retailer's credit card to judge whether the "discount" provided by the cash back offer is widely available or should be considered discriminatory. If deemed widely available, the CPI price should reflect the net purchase price (price less 2 percent cash back). If discriminatory, these offers should be excluded. In other cases, cash back offers are provided by the bank that issues a credit card. The cash back offer provides an incentive for consumers to use a particular credit card for purchases but does not represent a discounted price offered by the retailer and should be excluded. Cash back offers can also be tied to loyalty cards. A percentage of the total amount spent accrues and can be used as a discount on future purchases. These types of cash back offers are excluded because they apply to future purchases and the price paid is not discounted in any way.

(13) *Dual pricing for cash and for credit or debit card purchases*: some outlets may sell goods at different prices depending on whether the item is paid for in cash or if a bank debit or credit card is used. The primary objective of the price collection is to achieve a representative sample and continuity. The price collector should determine the proportions of buyers who pay by the different methods and this should be used to judge what prices should be collected to ensure representativeness. If the price collected on a particular month is based on cash payment, then a price based on

cash payment must be obtained in each of the following months. If a price collected on a particular month is based on the use of a bank debit or credit card, then the prices collected in the following months should be based on the use of a bank debit or credit card. Widely available reductions for cash payments may be included, but care should be taken to ensure representative purchases are measured, treatment is consistent from one period to the next, and practice adheres to the fixed-basket concept. Thus, where relevant, item descriptions should include method of payment and the proportions that are paid by different methods should be kept constant between reviews of the CPI basket.

(14) *Bank-specific credit card offers*: in some countries, banks that issue credit cards will negotiate with retailers to offer customers who use their credit cards a discount on all purchases. For example, a bank can negotiate with a supermarket to offer all customers using their credit card a discount off the total purchase price. Judgments need to be made whether such offers can be defined as generally available. Like discounts associated with loyalty cards, if the majority of consumers use that bank-issued credit card, it could be considered generally available and the discount reflected in the collected prices. However, if it is judged to be discriminatory and not generally available, they should be ignored.

(15) *Foreign currency denominated prices*: the price collector should collect prices of items quoted in foreign currency from outlets selling goods and services in foreign currency, if their exclusion from the CPI basket would undermine the representativeness of the index. The price should then be converted into local currency using an average exchange rate obtained by the head office from the central bank or appropriate dealers in the foreign-exchange market. The price in local currency (obtained after conversion) should be used in the CPI.

(16) *Product with a local and foreign currency price*: in some countries, there are outlets that sell products in local and foreign currency (that is, a purchaser is given an option to pay for an item in local or foreign currency). The local currency price is used when purchases are made in local currency. However, if more purchases are made in foreign currency, the foreign currency price is converted to local currency as in the previous example. To determine whether more purchases are made in local or in foreign currency, the enumerator must ask the outlet staff which currency is being used more frequently for purchasing the item, and this currency should be used for pricing purposes, unless purchases are frequently made in both currencies, in which case the price in both currencies is collected and a weighted average used after conversion to the local currency.

5.17 The appropriate practice to be followed will be determined by individual circumstances, which might vary among different countries. Price collectors should make extensive notes on the situations confronted so that decisions on how to treat specific cases can be quality assured and reviewed at the head office.

Unavoidable Costs That Are Not Part of the Advertised Price

5.18 In some cases, the consumer has no choice but to pay an extra cost, in addition to the advertised price of the product, to obtain the use of the product. In cases where most customers will treat and pay this cost as part of the purchase price, the extra cost should be added to the advertised price to determine the price for CPI purposes.

5.19 A common example of this situation is the sale of large household appliances and furniture. Most consumers would not be able to provide their own transport to carry these products home. As a result, many stores arrange delivery to the customer's home for an additional cost. It can be argued that most customers must pay for the delivery service, and this additional charge should be included as part of the purchase price for CPI purposes. The concept still applies when the transport is provided by another business, if this is part of the transaction. In this case, the transport cost would be priced separately but added to the cost of the appliance when the index is being compiled. However, if the customer arranges for the transportation under a separate contract then the transportation should be priced separately under transport services.

5.20 A similar situation can occur in marketplaces when live poultry is bought for meat. If consumers regularly buy live fowls and then go to other stalls to have the fowls slaughtered, then the two purchases can be combined to calculate the cost of buying poultry meat.

Price Bargaining[3]

5.21 Bargaining relates to a situation where prices are individually negotiated between sellers and purchasers and are not predetermined. In some marketplaces, the prices for a wide range of daily necessities are negotiated. Final transaction prices and quantities will vary from one transaction to another and cannot be determined until the purchase has been made. Similarly, there will be variations between transactions in the quality of the goods being purchased which will impact on the price that is paid. For example, the bargained price for fresh fruit is likely to be lower the longer the fruit is on display. Clearly, these special conditions require special methods to determine purchasers' prices for inclusion in the CPI.

5.22 Where prices are determined by bargaining, standard price survey methods, which consist of collecting prices directly from sellers, can generate erratic price indices that do not reflect actual price movements in a market. The prices that are collected should be "bargained prices" and will depend on the ability, willingness, and power of price collectors to bargain in the same way as genuine purchasers. In addition, prices can vary during the day and from

---

[3]It can be argued from the viewpoint of the *System of National Accounts* that bargaining is a form of price discrimination. A purchaser is not free to choose the purchase price because the seller can charge different prices for identical goods and services sold under the same circumstances. It follows that "identical" products sold at different prices should be recognized as having the same quality, and their prices must be averaged to obtain a single price relative to calculate price indices. In practice, the variation in transaction price can rarely be associated with identifiable price-related categories of customers. Rather, purchasers may inadvertently buy at a higher price than may be found elsewhere or could have been negotiated.

81

one day to the next, adding an extra dimension to the concept of representativeness. A number of survey methods and price collection techniques have been developed to overcome the difficulties inherent in measuring prices that have been bargained.

(1) *Survey by purchase of products, simulating a consumer*. The principle is that price collection should be carried out in conditions that simulate as closely as possible situations in which real transactions take place. Price collectors behave like regular consumers by actually purchasing items to be priced and spreading their purchases over the day to ensure representativeness. In each case, the field manager will need to carry out regular checks on quantities and prices obtained by collectors. The following approaches may be taken:

(1.1) Price collectors buy items to determine the relevant price through bargaining. They should be trained to behave as normal purchasers and strive to get the lowest possible price from selected outlets and sellers. Given the high turnover of sellers in many informal markets, the sample of sellers should be partially renewed on a regular basis to ensure that it remains representative.

(1.2) Price collectors buy items and, in addition, are given an incentive to get the best price. For example, a price ceiling may be set, and the collector may receive a proportion of the difference between the ceiling and the bargained price. This incentive system guards against potential difficulties caused by the collectors not getting the lowest price because, unlike an ordinary customer, they are not concerned with maximizing value for money and are not constrained by income although there might be a limit on the money available for purchasing.

(2) *Survey of purchasers*. The prices purchasers have paid are collected throughout the day immediately after the purchaser leaves the outlet or market stall, together with a record of the quantity and quality of the product purchased. The extent of the haggling or bargaining should be determined (for example, opening and closing prices) together with an indication of the relevant parameters determining the price. A form of incentive payment for survey participation may be needed where there is reluctance among purchasers to submit to such time-consuming questions. Determining the quantity and quality of the items purchased can be difficult.

(2.1) For the survey by purchase of products and the survey of purchasers, all CPI basket items subject to bargaining should be covered. The number of prices collected needs to be sufficient to cover all relevant items and to provide a reliable guide to average price. This may be difficult to determine beforehand, although previous price collections should provide some guidance. It is suggested that price collectors engaging in a survey of purchasers are given a form on which to record the number of quotations per stall or shop, as indicated by the various respondents. This can be used to check the number of quotes obtained against the target number set by the head office.

(3) *Survey of trends in wholesale prices*. A limited parallel collection of wholesale prices can be a useful supplement for items where the information obtained from the previous survey techniques is only partially successful, for example, where there is a deficit in the number of observations obtained. Ideally, prices should be obtained from the wholesalers where the relevant retailers get their goods. All factors should be observed that might result in increases in the corresponding retail prices, such as changes in taxes on retail activities, license fees, and the rental for the market stall. On the assumption that these factors remain constant over time, the evolution of wholesale prices may be used as a proxy for the retail prices of relevant items. The price of an item for the current period would be estimated by multiplying the price of the previous period by the corresponding evolution in wholesale price.

5.23 Determination of the prices paid by a purchaser can be challenging where the final price is for a bundle of items, for example, where a stall holder gives the purchaser extra quantities as a bonus for buying a number of goods. If the bonus comprises several categories of items, including the item on which a transaction price was being directly negotiated, then the purchase has to be split into as many subtransactions as item categories. In these cases, a common-sense approach is needed regarding treatment in the CPI. For example, consumers who are living on a subsistence income consume all food purchases and the extra quantities involved will be consumed and should be included when calculating the price paid. Purchasers will have actively bartered an overall price for the total basket of purchases, including any "free" goods thrown in.

5.24 The method for determining the price paid by the purchaser is illustrated in the following example: a purchaser wants to buy five kilograms of carrots and is offered a bonus consisting of 500 grams of carrots, a head of lettuce, and three onions.

5.25 Three transactions can be identified, involving: 5.5 kilograms of carrots, 100 grams of lettuce, and 200 grams of onions. The bonus should be valued at prices at which the seller would have sold the items and the purchaser would have bought them. The assumption made is that prices would have been determined through bargaining on the same conditions as the price of the item that initiated the purchase (carrots). In this example, the opening value of five kilograms of carrots is 15,000 pesos and the closing value 12,000 pesos, whereas the opening values of other food products included in the bonus are 990 pesos for a bunch of 264 grams of lettuce and 4,620 pesos for a pile of onions of 4.4 kilograms. The actual closing price of carrots will be determined as shown in Table 5.1. The actual purchaser's price of carrots is found to be 2.0967 pesos per gram or 2,096.7 pesos per kilo.

5.26 If the price collector does not know the closing price at which lettuce and onions would have been sold by the seller of the carrots, it can be estimated. This is done by collecting opening values and standard quantities from a

Table 5.1 Determining Purchaser Price When Bargaining Takes Place

		Requested Item	Bonus Items		
		Carrots	Carrots	Lettuce	Onions
(a)	**Opening Value of Standard Quantities (local currency units)**	15,000	15,000	990	4,620
(b)	**Standard Quantities (grams)**	5,000	5,000	264	4,400
(c)	**Opening Unit Price of Standard Quantities (local currency units per gram) (a/b)**	3.00	3.00	3.75	1.05
(d)	**Opening Unit Price of Quantities and Bonus Quantities (local currency units per gram) (as per (c))**	3.00	3.00	3.75	1.05
(e)	**Quantities and Bonus Quantities Offered (grams)**	5,000	500	100	200
(f)	**Opening Value of Quantities and Bonus Quantities (local currency units) ((e) at unit price at (d))**	15,000	1,500	375	210
(g)	**Closing Value of Items Received (local currency units) (assumes bargaining ratio for lettuce and onions same as for carrots, that is, 15,000/12,000 = 1.25)**	12,000	1,200	300	168
(h)	**New Price (local currency units per gram) (g/e)**	2.40	2.40	3.00	0.84
(i)	**Bargaining Ratio (D/H) (for example, old price for carrots/ new price for carrots, that is, 15,000/12,000 = 1.25)**	1.25	1.25	1.25	1.25
(j)	**Payment (local currency units)**	12,000			
(k)	**Estimated Closing Value of Bonus (local currency units) (1,200 + 300 + 168 = 1,668)**	1,668			
(l)	**Actual Value of Requested Item (all carrots) (local currency units) (12,000 − 1,668 = 10,332)**	10,332			
(m)	**Quantity Received of Requested Item (grams)**	5,500			
(n)	**Actual Purchaser's Unit Price of Requested Item (local currency units per gram) ((12,000 − 300 − 168)/5,500) = 2.0967)**	2.10			
(o)	**Improved Bargaining Ratio (d/n), that is, 3/2.10 = 1.43**	1.43			

sample of sellers in the same market or at different outlets in the same area. The average opening price of an item is equal to the sum of opening values of the item divided by the sum of relevant standard quantities. For each bonus item (lettuce and onions), the resulting average opening price will be divided by the bargaining ratio calculated on the item needed (carrots) to estimate a closing price for that bonus item. The value of each bonus item is obtained by multiplying the closing price by the quantity offered. If the packet of bonus items contains an item of the same quality as the requested item, that bonus item will be valued based on the closing value of the requested item.

The Principle of the Fixed Basket

5.27   The important principle underlying price collection is the need to compare prices on a like-to-like basis from one period to the next. This has consequences:

(1) Where there is a choice of variety of product to be priced initially, an important consideration should be whether that variety will be available to price over a reasonably long period. This is in addition to being typical of what is sold to customers. Note that tight specifications are of no use if the described item cannot be found in the outlets.

(2) A record should be kept of additional information needed to ensure the unique identification of the variety priced so that:

(2.1) The same variety continues to be priced in the case of subsequent price collection being carried out by a different person.

(2.2) The identification and adjustment for any quality change can be made when the variety disappears and is replaced by a different one.

Item Specifications

5.28   The recording of item specifications is particularly relevant for traditional forms of price collection where products are sampled and where price collection involves price collectors visiting retail outlets to collect prices or where prices are collected by postal inquiry, email, or telephone most particularly from central sources.

5.29   There are no firm rules, especially regarding the use of loose or tight item specifications: each NSO may choose their preferred methods. Considerations in deciding on item specifications are the following:

(1) Loose specifications leave more discretion to the price collector, so the reliability and training of collectors are factors to consider when deciding whether to use loose or tight specifications.

(2) The specification should be more detailed for heterogeneous items where there is scope for significant difference of varieties, and for items that by nature are highly specified, such as cars and hi-tech goods.

(3) Loose specifications allow the index to more broadly reflect regional differences in tastes and preferences, and account for differences resulting from socioeconomic factors.

(4) Tight specifications allow for the calculation of meaningful average prices.

(5) Average prices are useful to identify outliers and to assess the accuracy of reported prices.

(6) Subject to satisfactory sample design, average prices allow comparisons of price levels, including between regions or urban and rural areas. Tight specifications facilitate the use of CPI prices in the computation of purchasing power parities.

(7) However, specifications should not be so detailed that the item cannot be found in an adequate number of out-

83

lets. Individual varieties sold differ from outlet to outlet and region to region, and the CPI should reflect these differences. Overly tight specifications do not reflect this diversity and can increase the number of missing prices.

(8) Without tight specifications, average prices have little meaning or use. For example, an average price for a liter of whole milk is more meaningful than an average price for men's shirts.

5.30 Responsibility for deciding whether to use loose or tight specifications and for specifying the items to be priced should normally rest with the head office. In general, loose specifications can be useful for food, beverages, clothing, and personal items. Tight specifications tend to be useful for electronic and other items with high rates of turnover. Specifications should be reviewed on a regular basis to determine whether they continue to be relevant. The following could signal the need to review and revise specifications:

- A large number of missing price quotations
- A large number of substitutions
- A wide variation in the distribution of collected price levels
- An observed distortion in the achieved sample, for example, where an item is stocked only by one retail chain indicating that the item is unlikely to be fully representative of what households buy, or where there is a reduction in the number of brands available possibly indicating a fall in demand

5.31 Practical difficulties stem from the temporal dimension in price collection. Issues such as the timing of entering prices into the index, the treatment of missing prices, seasonality, and the frequency of price collection are temporal issues that need to be addressed to maintain the quality of the index. Aspects relating to the frequency and timing of price collection are addressed in paragraphs 5.35–5.47.

Frequency

5.32 Decisions about the frequency of price collection are governed by several factors. The most important are: the volatility of prices; the characteristics of the market from which the prices are collected; the known regularity of price changes; and the frequency and method of calculation of the CPI. The general principle is that each item should be priced as often as is necessary to ensure that the index reflects a reliable and meaningful measure of price change. Thus, the desired frequency of price collection will vary by product, depending on how frequently the prices to be observed change. For instance, the prices of some products (for example, fees for government services and utilities) might need to be collected only once a year if it is known that prices are reviewed annually at a regular point in time. However, it is advisable for the price collector to occasionally check that the assumption of regular and predictable reviews still holds. In contrast, the prices of products with more volatile prices, such as fresh food, might have to be collected more frequently than the frequency of index calculation and publication. Also, price collection will need to reflect user needs with regard to whether the target index is a point-in-time or period-of-time index.

Period-of-Time or Point-in-Time Price Collection

5.33 Prices for a monthly (quarterly) CPI should aim to reflect the average price of the reference period. To that end, NSOs should strive to calculate an index based on prices covering the whole period (for example, a month). Ideally, price collection would be organized so that prices are collected from different outlets throughout the month. This ensures that the prices used to calculate the index more broadly reflect the average price for the reference month. While resource constraints could restrict price collection to a specific period during the month, effort should be made to maximize the price collection period to include as many days in the month as possible. Regardless of the price collection period, the interval between successive price observations must be held constant, for example, by visiting an outlet during the same fixed time period each month or quarter. If the CPI is used to deflate income, expenditure, or sales, the index should relate to the time period of these money flows. For economic analysis, where the index will be used in conjunction with other economic statistics, most of which relate to a period rather than to a point in time, the CPI should do the same.

5.34 Under the traditional approaches to price measurement, where price collectors visit retail outlets and record prices, spreading price collection over a period of time will result in a more even workload. This avoids some of the operational problems and inefficiencies associated with point-in-time price collection but will involve collecting sufficient price quotations to obtain a reliable average price level for each product over the period.

5.35 From an operational perspective, the uneven workloads associated with point-in-time price collection when using traditional price collection techniques can be inefficient and have a negative impact on price collector performance at peak times. For instance, a sizeable price collection team will be required for a short time in each collection period. This has implications for the recruitment and deployment of price collectors and for management of the fieldwork. Communication and planning between collectors and their supervisors to manage events like absence because of illness must be prompt and effective. Staff at the head office will similarly be confronted with a heavy workload of data checking and editing of price data over a short period of time.

5.36 With the point-in-time approach, major price setters, notably the government, can influence the index according to whether their price changes take effect on a day just before or just after the day for which their price information is obtained, or on the day of collection. Since prices are often collected centrally from such price setters, it should be possible to obtain information from them about the amount and timing of price changes at the end of each month, so that in applying the period-of-time approach, an average price for the whole month can be calculated. For example, if electricity charges are made quarterly and prices increase partway through the three-month period, individual customers' payments could include zero, one, two, or three months at the higher rate depending on when the price change is implemented.

5.37 The point in time or period is chosen to represent a prespecified reference period, normally the calendar month

in which the price observations take place. Whether collection is continuous or point-in-time, the interval between successive price observations must be held constant, for example, by visiting an outlet during the same fixed time period each month or quarter.

5.38 The sampling variance will differ according to whether a period or point-in-time index is compiled and the frequency of collection. Regarding the timing and frequency of price collection, the CPI compiler should consider the trade-off between statistical accuracy and cost, particularly when prices are collected by visits to local outlets, as this can be a costly activity. The budget for price collection can limit the available options.

5.39 The timing of the publication of the resulting price indices can be a constraint on the price collection schedule and vice versa. For example, there may be legal constraints on the timing of the publication of indices, such as a requirement to publish at a set time each month. In such cases, prices must be collected to a schedule that allows quality assurance, processing, and aggregation procedures to be completed before the deadline. The quality management of the price collection process is covered in paragraphs 5.78–5.116.

Timing of Price Collection

5.40 The interval between price observations should be uniform for each outlet. Since the length of the month varies, this uniformity must be defined carefully, for instance, not by date but by a formula such as "second Monday in the month."

5.41 Regular timing is important, particularly when inflation is high. Where there is a specific collection day, the most volatile prices should be collected on that specific day rather than on the days around it. Items where prices can be more volatile include fresh fruit and vegetables, fresh meat and fish, and fuels. In the case of food products sold in marketplaces, the time of day as well as the day of the week matters. Prices of fresh fruits, vegetables, meat, and fish can be higher in the mornings when produce is fresh and lower in the evenings, especially if there is limited refrigeration. Thus, the precise timing of price collection is particularly important for fresh produce.

5.42 When inflation is low and stable there will be little difference between the annual inflation rates based on point-in-time and period-of-time collections. For example, there is likely to be very little difference between the annual rate of change in the index from Monday, January 8, 2018 to Monday, January 7, 2019 and the corresponding annual rate of change between the complete months of January 2018 and January 2019. This will not be the case if inflation is high or the inflation rate changes significantly during the year. The difference between January 1 and February 1 and between average January and average February inflation rates may be different, particularly if so-called "sale" periods fall on regular dates each year or are limited by laws. For certain products with high index weights, where price changes are sudden and tend to affect the whole market on about the same day, the choice between point-in-time or period-of-time price collection is especially influential. Examples are fuels, electricity, and telecommunication prices. For these kinds of products, it can be argued that the case for taking an average price for the period is stronger. This ensures a more

meaningful measure of price change for the month. The calculation of the average monthly price should relate to the periodicity of the collection, taking into account the appropriate pricing periods. For example, if prices are increased a third of the way through the period, then two-thirds of the average price over the month should reflect the higher pricing. In these cases, different areas or locations should be scheduled for price collection at different times of the month according to a regular pattern to be repeated each month. This makes the use of the collectors' time more efficient and has the advantage of providing a spread of collection dates for many representative items. Individual price observations should be carried out at the same time each month so that the index does not move because of a change in the length of interval between collection dates. Prices may vary by day of the week (for example, depending when is market day) or by time of the day (for example, fish is more expensive in the morning when it is fresher), and prices should be collected on the same day, at the same time, if this is the case.

5.43 Preferably, days of the week and month should be chosen taking into account when purchases are concentrated and where prices and goods in stock are known to be representative of the whole month. In some countries, the results of the household budget survey suggest that most households do the shopping on the day of the market or souk. However, retailers may be less prepared to cooperate when they are busy, so a balance needs to be struck between the ideal timing for collection and the impact on response rates.[4] An entirely fixed interval is impossible because of the varying length of a month and the timing of public or religious holidays. One solution is to take sequences of four and five weeks, so maintaining a relatively stable monthly or quarterly period where price collection takes place on a fixed day or days each month; another is to follow a rule such as collecting prices on the regular market day or on a Wednesday through Friday of the first full week in the month.

5.44 Price collection days (and sometimes the time of collection) need to be set in advance. The NSO should explain the procedures used for setting collection dates and the underlying objectivity of the method to assure the public of the integrity of the index. Any data suppliers who supply prices directly to head office staff need to know the collection date in advance to be able to prepare and supply the necessary price. To enhance transparency, it is recommended practice to include the price collection period in the CPI metadata.

Measuring Hyperinflation or Selected Large Price Changes

5.45 Special arrangements need to be put into place in case of hyperinflation. Hyperinflation is an economic phenomenon that occurs when inflation increases very rapidly. Economists typically consider monthly inflation rates of above 50 percent as hyperinflation episodes. In these circumstances, it becomes even more important that the prices

---

[4]There is no rule of thumb on what constitutes an acceptable response rate as this will depend on the sample design that has been adopted and the structure of the retail sector (most particularly, the variability of product and item varieties available and their prices, and the range of outlets in which they are stocked).

of individual items in individual outlets are collected at precisely the same time each month, otherwise misleading figures may result. The NSO should consider collecting prices more frequently and, correspondingly, a more frequent compilation of the index. For example, where prices are normally collected on a quarterly basis, it may be sensible to collect them more frequently. If this is not feasible, it may be appropriate to adjust prices proportionately by some relevant indicator, such as a subset of the index, to provide an approximation to a monthly index. An appropriate comparator must be chosen because relative prices can change dramatically in periods of hyperinflation. The same considerations apply to collecting prices online.

5.46　In some circumstances, rapid or frequent price changes may be associated with certain items only and action should be taken accordingly. For example, food prices may rise disproportionately because of a bad harvest and it may be sensible to increase the frequency of the index for food items only, possibly publishing a separate index. Alternatively, a simpler way of dealing with this situation may be to monitor a small number of relevant prices on a regular basis without producing a full price index. Such subindices can be published separately or used to adjust the index as mentioned previously, or to provide background briefing for analytical purposes. These items may be chosen according to their importance for the household budget and whether they are particularly susceptible to big price increases.

The Practical Aspects of Managing Price Collection

5.47　As mentioned in the introduction, the most appropriate sampling and survey procedures will vary depending on the use of the CPI and on national circumstances. This also applies to the practical aspects of price collection. Paragraphs 5.51–5.116 describe aspects of planning and organization and provide indicative guidance on the processes and procedures that contribute to successful price collection in the field.

Practical Procedures for Local Price Collection: Planning and Organization

5.48　The discussion that follows is based on period-of-time pricing rather than point-in-time pricing (see paragraphs 5.36–5.42), but the concepts discussed generally apply to both collection methods.

5.49　The procedures governing the collection of prices have the requirements of obtaining usable prices from the outlets, and the practical problems faced in managing travel to and from the various locations, in transferring data, and in validating the data back in the office. The overall operation can only be achieved by cooperation between the price collectors, their supervisors, and, of course, the retailers selected for the price survey.

5.50　An overview of local price collection is given in Annex 5.1. It is in the form of a flow chart and shows the different situations that confront the price collector and the decisions or referrals to the head office for further action.

Planning the Collection Schedule
5.51　Price collection is a complex process that needs to be effectively managed by the head office. Proper training

of price collectors in the concepts and practices of price collection is an essential element in achieving a representative and error-free sample of prices. Price collectors should be given help and practical support when undertaking price collection. The collection of prices should be supported by good documentation. Also, the scheduling of price collection and the drawing up of contingency plans, for example, in the case of price collector illness or other factors outside the control of the price collection team should be in place so that the price collection operation runs smoothly. Effective management of the price collection process should be given the necessary attention and funding. More detailed guidance on good practices is given in the paragraphs that follow.

5.52　The collection schedule should include time for the price collector to travel around all required locations within a reasonable number of working hours in a day. The schedule needs to allow the price collector to perform all the necessary checking of prices, to answer queries from the supervisor or the head office, and where necessary to revisit the outlet. The collection schedule also needs to allow for the transfer of data and forms between the data collectors, regional offices (if applicable), and the head office. If paper forms are being posted or hand-delivered, time must be allowed to ensure that all information reaches its destination by the required deadline. If electronic transmission of data and forms is used, time needs to be allowed to ensure that all the data arrives in the correct format and to address potential problems if the files are corrupted. Plans for recovery in case of technical breakdowns need to be in place.

5.53　Public holidays can occur on days that would otherwise be price collection days. In general, prices due to be collected on the holiday should be collected as close as practicable to that day. This usually means adjusting the regular price collection schedule to the immediately preceding or following few days. Any adjustments to the regular schedule should be made so that prices reflect the normal pattern of buyer–seller behavior. The work schedule for index compilation and publication also needs to be considered.

Dealing with Queries: Inquiry Management System
5.54　Price collection queries need to be dealt with in a timely and efficient manner both because of the tight schedules associated with the compilation of a CPI and because of the difficulty to correct errors in pricing retrospectively given the dynamic nature of retailing where prices and stock can change very quickly without warning. The progress in dealing with queries (for example, verify questionable prices or seek additional details on incomplete specifications) needs to be monitored and the system for monitoring must be simple, in order to be effective and flexible to the needs of the CPI production cycle and to any problems that may arise. An inquiry management system should be able to monitor progress and provide an effective audit trail. Information recorded should include: date price collection form received and from whom; date due; collection date/trip/period (if applicable); current progress (with date); dates when queries were raised and responses sent; and date completed. It is particularly important that decisions are recorded and signed off to provide price collectors with

confirmation for their own purposes of verification and as audit trails for quality assurance purposes at the head office. Audit trails are an effective way to ensure that processes were carried out correctly and to support the review of the effectiveness of those processes. The information gathered should also feed into a quality management system for the computation of the CPI (see Chapter 13).

5.55 A simple inquiry management system might use emails to send the query. As these will be seen and read, they satisfy the notification requirement. A template form requiring the entry of simple ticks and dates could be created in a spreadsheet and a new copy provided for every collector each collection period.

5.56 An inquiry management system can be either paper or electronic. The two methods should be matched to the resources and infrastructure available. For instance, messages to and from price collectors to their supervisor might be on paper but messages between a regional office and the head office could be by email.

5.57 Queries about collected prices are a two-way process. The head office may want to question some prices collected and request a check, while price collectors may ask for further guidance on situations arising in the field such as the selection of a suitable replacement for a disappearing good.

Practical Collection Procedures and Questionnaire Design

Design of Price Collection Questionnaire

5.58 Good design of the questionnaire form (or its electronic equivalent) is essential for the successful collection of prices (see Annex 5.2). Price collectors should be given the appropriate direction and find it easy to use, and the format and layout should facilitate the extraction of data (for example, price, item description, or comments) by the head office for effective quality assurance. The price collection form should cover: collection date and collector's name; name or location of the particular outlet; product name and specification of the actual item to be priced; price entry; and collector comments about the product or price movement or changes in the representative item being priced such as package size.

5.59 Whether recording price information obtained from personal visits on paper form or via handheld computer or tablet, the price collector needs to be provided with all the information required to successfully collect prices and, correspondingly, the head office will want from price collectors all the information needed to assure the quality of the prices collected.

5.60 Including the previous period price on the collection form continues to be a topic of debate. There are advantages and disadvantages to including the previous period price on the collection form. The optional inclusion in the form of the price collected in the previous period raises concerns that the price collector may be inclined to automatically record the previous price or be too influenced by it when identifying the item in the outlet to be priced. However, this information is sometimes included to assist the collector in ensuring that the correct item and price are being recorded and, where paper forms are being used, to identify any unusual price movements that need to be investigated. With electronic collection, if the collected price differs from the previous period price, data collectors are prompted to justify or explain the price change. Price collectors are encouraged to add additional information to the "standard" description to facilitate the unique identification of the product to be priced without a reliance on price. Such information might include: brand, make, size, model name and number, reference number, distinctive features, and location in outlet (for example, "bottom shelf at rear of shop").

5.61 Items should appear on the form in the same order in which prices are collected. A correctly ordered list will reduce time-consuming swapping and searching between pages. Even with a handheld computer or tablet, searching and pressing of navigation buttons will be less time-consuming if the item list is ordered to match the layout of the outlet.

5.62 For fruit and vegetables, weight and quantity should be part of the item description, but the amount actually priced should be recorded separately so that a unit price can be calculated. Weight and quantity are needed because the unit price is often lower if larger quantities are purchased. For example, if the price collector prices a bunch of bananas and the item description is one kilo of bananas, the price collector should choose a bunch weighing approximately one kilo, weigh the bunch, and calculate the price per kilo.

5.63 When an index basket is updated, the questionnaire will need to list all items included in the old and the new basket. Where an update of the outlet sample is planned to coincide with the introduction of a new basket, it will be necessary in the chain link period for the price collection to cover both the old sample of locations, outlets, and items and the new sample of locations, outlets, and items.

5.64 An example of a typical price collection form is given in Annex 5.2. This example is of a form used by the collector for recording prices when visiting an outlet, either in paper form or electronic. It is also possible to ask the outlet staff concerned to complete the form themselves and to send it to the NSO, although there are obvious potential problems in the form not being completed correctly. Such a form may therefore serve for reporting as well as for collecting. If the form has space for recording prices over a sequence of months, the collector may keep the form and transcribe the prices each month onto a separate report form, which is then sent to the NSO. Where the form used for collection is also used for reporting, there are two main options: either the form has space for recording prices over a whole sequence of months, and the form is shuttled backward and forward monthly between the collector and the head office, or new forms for collection and reporting are provided each month. In the latter case, it can be useful if the form contains the prices recorded in the previous month alongside the spaces for recording the current month's prices, although this might encourage the price collector or data provider to overly rely on the previously recorded price. The transfer of the prices to another form or system, especially when done manually, may lead to transcription errors and is best avoided. In the example given in Annex 5.2, the previous month's price is shown (see paragraphs 5.185–5.193 on Computer-Assisted Data Collection for a more detailed discussion of the advantages and disadvantages of giving price history). For simplicity, the example assumes all prices entered on the form will include a sales tax.

87

5.65   Completion of the price collection forms, whether by paper or electronically, should comply with the following guidelines:

(1) All prices should be entered in the collection sheet in full even if there is no price change. This is good practice as it helps to ensure that the correct item is being priced, the price has been recorded correctly, and the price collector has not relied on information given by the outlet staff without checking. For instance, a price may not have changed but the package size might have been reduced, a practice sometimes used to make price increases less transparent, but to the outlet staff, there has been no price change. A price should never be recorded as "no change." All other entry fields on the price collection sheet should have something entered if only to indicate that they were not accidentally missed, for instance, "Not Available (NA)."

(2) If a price is not available a reason must always be supplied. The information given will be useful to the collector's supervisor and the index compilers as well as to the price collector. The price collector may need to ask the outlet staff for an explanation and may also need to choose a replacement, for example, if the outlet staff indicates that the item is no longer stocked.

5.66   No matter which method of local price collection is used, it is essential to provide some procedures to allow the tracking of activities and formal sign-off, confirming that processes have been completed and necessary actions undertaken on data checking and transfer to the head office. Audit trails are essential for work and quality management.

5.67   The price collection form should have space to provide full descriptions of the items being priced. Price collectors should be given a checklist or set of codes to record relevant information on changes relating to outlets, items, or prices. The information needs to be systematically collected. For instance, codes to help with quality adjustment need to reflect those characteristics that most influence price. Prior research, for example, based on the hedonic method (see Chapter 6), can help to predetermine these characteristics.

5.68   Codes for managing the sample of outlets can be helpful and may include:

- *Closed down*: outlet permanently shut or closed down
- *Temporarily unavailable*: outlet temporarily closed, but likely to be open next month
- *Refusal*: owner or staff refuses to cooperate
- *Change of details*: change of ownership or name or change of purpose

Continuity and the Use of Price Collection Codes

5.69   As the index measures pure price changes, the same item should be priced every month to establish a true picture of price change. For example, if a jar of a supermarket's own brand strawberry jam has been selected, that brand and flavor should continue to be collected; if it is out of stock, another brand and flavor should not be used without further investigation to establish whether this is a temporary situation or likely to be permanent. In the latter case, and if another flavor of the same brand, size, and quality is

available, then in normal circumstances, this item should be chosen as a "comparable" item and the item description suitably amended.[5] If a different brand, size, or quality product is available then this should be selected as a "new" item, but only where no comparable items are available. The same principles apply to other items such as clothes and fresh fruit and vegetables. With clothes, it is important that color, fabric, country of origin, logos, and size are specified to ensure that the same item is priced each month. For fresh fruit and vegetables, useful attributes to record may be country of origin, class, and variety. For electrical equipment, the specifications and features given in the manufacturer's catalog may be important.

5.70   A detailed description of the items being priced is recorded to assist the price collector and the head office in choosing or confirming the suitability of a replacement for an item that has been withdrawn and in identifying changes in quality. The focus should be on recording price-determining characteristics. If the regular price collector is unable to carry out the normal collection, full and accurate descriptions will also enable a relief collector to carry out the collection without any doubt concerning the correct items to be priced.

Most of the time, the item will be exactly as collected in the previous month, and all that will be recorded is a current price. However, if there is a change or uncertainty in the item, it is necessary for price collectors to use their judgment and to inform the head office, bearing in mind that the head office staff is responsible for making the final decision. A precoded specification is less time-consuming and provides better guidance to the price collector on what information should be reported. The codes should cover situations faced regularly when collecting prices. The codes can be numeric or alphanumeric, and each one should indicate the action taken by the price collector or supervisor and will be associated with corresponding procedures to be followed in the calculation of the index by the head office. The form should include codes such as:

- *Comparable (C)*. The original item is no longer stocked but a similar alternative has been collected that does not differ with regard to major price-determining attributes. The price is likely to be in a similar range although this may not always be the case.
- *Noncomparable (N)*. The original item has been replaced by a new item that is not really comparable but is equally representative of that product group. If possible, the collector should try to find the price of the "new" item in the previous or price reference period.
- *Sale or special offer (S)*. A price decrease because of a genuine sale or special offer. This does not include damaged or out-of-date stock or clearance goods, which should never be included. A price reduction where there is no notice of a sale or special offer is not a "sale"; the item should still be priced, but without the S indicator code.

---

[5] The replacement should be either (1) as similar as possible to the previous one; (2) the most popular "similar" product in the outlet; or (3) the "similar" product that most likely will be available for future pricing. Unlike approach (1), which leaves the sample "static" with the risk that it will be increasingly out of date and difficult to collect prices for, approaches (2) and (3) have the advantage of introducing an element of sample replenishment.

88

- *Recovery (R)*. A return to the normal selling price after a sale or special offer. This does not need to be a return to the price prior to the sale or special offer. A referral back to the previous price collected and consultation with the outlet staff will normally be required.

- *Temporarily out of stock (T)*. Guidance will need to be given to the price collector concerning the meaning of "temporarily" (with regard to expected duration, which may vary for different items). If nonseasonal items are missing for three or more months, depending upon country circumstances, enumerators should replace items immediately (for example, fashion clothing, if it is unlikely that the identical item will come back into stock). Typically, "T" indicators should not be used for more than three consecutive months, and in the fourth month, a replacement should be chosen. In food outlets, it is very unusual for items to go permanently out of stock. The collector should always try to check future availability with the retailer.

- *Missing (M)*. Used where the outlet does not stock an item or no longer intends to stock an item and there is no appropriate alternative replacement. In these circumstances, it is recommended that the item is checked at subsequent collections to ensure that a suitable replacement item has not come into stock.

- *Weight (W)*. A permanent weight or quantity change to the product.

- *Query (Q)*. Such a code may be used to supply extra retail information to the head office (some examples are "10 percent extra free," "2 for the price of 1," or a strange price difference that is not covered by one of the other indicators, such as a special edition issue of a magazine at an increased price). Arrangements need to be in place for the head office to respond to these comments and to treat the price quotes accordingly.

5.71   Even if the retailer says there have been no price changes since the previous month, the price collector should confirm prices. This will require some diplomacy, but it is important because the outlet staff may overlook a small number of price increases, or forget when the last increase has occurred. Also, checks need to be made to ensure that there have been no changes in the price-determining characteristics of the item such as package size or weight.

5.72   As noted previously, a price should be recorded only if the exact product being priced is on display and immediately available for sale; however, for certain large items that must normally be ordered, such as furniture or cars, the price should be recorded as long as the retailer confirms that it is available for delivery.

Unit Prices

5.73   Some food items, such as meat, fish, or cheese, are typically sold in variable weights and it is necessary to collect prices per unit of weight. The price per unit of weight should be taken from the package labeling or calculated directly by the collector. Roughly, the same package size and type should be used each month, as the unit price might be lower for larger pack sizes or differ between package types. Other items, such as eggs, are often sold in specified quantities. For these, collectors record prices for the specified quantity, as total and unit prices usually depend on the number bought. If X eggs are to be priced and the price for the number is not quoted directly, the price of one egg can be obtained and multiplied by X to get the required price. However, the price collector must ensure that the unit price does not decrease with quantity: significant changes in weight or quantity are to be avoided. Other examples are herbs, such as mint, or vegetables, such as cabbage leaves, sold in bunches of variable size with no label giving the weight. In this instance, a number of bunches should be weighed and priced to obtain an average price per unit of weight (for example, kilogram). It is necessary to purchase bundles and weigh them at the head office when reliable scales are not available at the outlet or market stall.

5.74   Certain food items, such as fruit or vegetables, are more difficult to price as some outlets might price items per number purchased while others might price by weight. For example, peppers may be priced either by weight or by unit no matter what the size; garlic may be priced per bulb, clove, or by weight; and various types of berries may be priced either by weight or by container, which may differ in size or how full it is. In these instances, care needs to be taken with the product descriptions. The item description given on the price collection form should refer to the pack size, weight, or quantity being priced. Collectors need to be aware of the importance of collecting the same item from one month to another, so that genuine price changes are recorded and not quantity or quality price changes.

Practical Collection Procedures: Quality Management in the Field

5.75   Adequate field procedures are required to ensure that the quality of the price index is not compromised by errors in price collection. Price collection needs to be carefully planned and managed, and effective instructions and training are given to price collectors. In traditional price collection, most prices are collected through price collectors visiting individual outlets. Checks need to be carried out to ensure the accuracy of the data. This section gives guidance on field procedures relating to local price collection and provides an overview of quality management. It focuses on data validation in the field. Chapter 13 provides a broader look at the organization and management issues associated with central price collection and with the complete process of producing a CPI.

Data Validation in the Field

5.76   Data validation should be carried out throughout the entire process of compilation of a CPI, from the collection of individual prices to their aggregation into indices.

5.77   The questionnaires, and the data collection software if handheld computers or tablets are used, should facilitate quality assurance of the collected prices at or close to the time of collection and record the results of the checks made in the field as part of an audit trail.

5.78   Collected prices can be compared to the prices of the same product sold in the same outlet previously collected, and large movements can be checked for accuracy. Preferably, the head office should provide guidance on what are viewed as acceptable ranges for price movements based on previous price movements. If a price has changed

significantly or remained unchanged for a very long period, the retailer should be asked to explain why.

5.79   Where resources allow, field supervisors and independent auditors should be deployed to support price collectors in providing accurate prices for input into the CPI. Supervisors should check the validity of the prices and related information recorded by the price collectors, provide assistance, and help when required as part of a collaborative effort. The required level of the validity checks may vary depending on the nature of retailing and the data collection procedures. For example, the use of handheld computers and tablets facilitates real-time data editing and the creation of price collection reports, reducing the chance of errors in price collection.

Data Validation: Field Supervisors

5.80   Checks to ensure that data are complete and correct should be carried out as early as possible in the collection and compilation processes. A return to the outlet to re-input prices becomes increasingly less feasible as time goes on, and there is a greater risk that the prices in the outlets will have changed since the initial collection. The use of handheld computers or tablets by price collectors facilitates much more detailed checking at the time of the initial collection of prices in the outlet than the equivalent paper system (see paragraphs 5.185–5.193 on electronic reporting).

5.81   Field supervisors have a number of important roles: training of data collectors when implementing new procedures or methods; one-on-one training of price collectors during joint field trips to correct any deviations from the procedures and tasks laid down in the price collection documentation; and reviewing the work done by price collectors in previous days to verify the quality and to facilitate the correction of errors.

5.82   In an ideal system, field supervisors will be employed to regularly check that price collectors are adhering to the price collection schedule and are undertaking the required checks at the appropriate time. The supervisor should check that price collectors are completing price collection forms correctly. A sample of collection forms from each collector should be checked where it is not practical for supervisors to check every form. Checks may be made, for example, on whether the price collector has attempted to collect all prices from all outlets, that explanations have been given where prices were not obtained, and adequate descriptions entered where replacements for disappearing products have been priced. The supervisor may also be required to check the accuracy with which data are transferred from data collection forms to the computer. This is an essential task associated with the quality assurance process and needs to be allocated to somebody other than the person who initially inputs the data so that an independent check can be performed.

5.83   Supervisors should also be encouraged to visit outlets and check the individual prices collected by price collectors. These checks can be organized either on a random basis or chosen on the basis of indicative information, such as extreme price variations. A typical audit report will include the percentage error rate and a breakdown of whether the errors are likely to have a high impact (for example, wrong price, wrong item, or item available but listed as temporarily unavailable) or a low impact (for example, incomplete product description or inadequate reason given for price change). The audit report should be followed up by a formal request to the price collector asking for corrective action and confirmation that all necessary follow-up actions have been carried out.

5.84   Field supervisors should check for consistency and credibility in price movements recorded by the collectors under their oversight. For instance, if one location is reporting different price movements from the other locations within the collection region, some explanation or a follow-up price collection will be required to check the accuracy of the prices collected. Preferably, this should be done once data have been transferred into the computer system and have been checked for errors. Tabulations of price changes grouped by product or elementary aggregate should be provided to enable the supervisor to conduct these checks efficiently. This will enable the supervisor to quickly identify extreme or inconsistent movements, which may indicate either errors in collection or unexpected behavior in the market. These checks should be conducted regularly during the price collection period.

Quality Checks in the Head Office: Data Entry Queries and the Role of the Head Office

5.85   Once the price collection has been completed and the prices submitted to the head office, a series of further validation checks can be run. In determining the checks to use, the validation checks carried out in the field should be considered. For example, the use of handheld computers or tablets will increase the potential for validation at the time of price collection and reduce the need for detailed scrutiny at the head office. In addition, it would clearly not be productive or cost-effective to repeat tests already carried out.

5.86   The range of tests carried out on individual price quotes can include:

(1) *Price change.* The price entered is compared with the price for the same product in the same outlet in the previous month and triggers a query where the price difference is outside preset percentage limits. These limits will vary, depending on the item or group of items, and may be determined by analyzing historical evidence of price variation for the product or item concerned. If there is no valid price for the previous month because, for example, the item was out of stock, the check can be made against the price two months or three months before.

(2) *Maximum/minimum prices.* A query is raised if the price entered exceeds a maximum or is below a minimum price for the item of which the particular variety is representative. The range may be derived from the validated maximum and minimum values observed for that item in the previous month expanded by a standard scaling factor. This factor may vary between items, again based on previous experience. Where necessary and possible, the maximum/minimum price should take account of any significant differences in average prices between, for example, regions.

5.87   If a handheld computer or tablet is used (see paragraphs 5.185–5.193), both tests (related to price change and maximum/minimum prices) can be implemented easily to

90

take place at the time of collection; otherwise, they will need to be conducted in the head office as soon as possible after collection and prior to the computation of the index. A failure in either test should prompt the collector to check and correct or confirm the entry, and prompt for an explanatory comment.

5.88   Queries raised may be either dealt with at the head office or sent to the price collector for resolution. For example, scrutiny of a form might show that a significant price difference has arisen because the item priced was a new product replacing another that has been discontinued. In this case, there may be no need to raise a query with the price collector, unless there is evidence to suggest that labeling the item "new product" is incorrect.

5.89   If a price collection error is discovered and is too late in the process of computing a nonrevisable CPI for the collection of the correct price, the head office will need to reject the price and exclude that item from that month's index and the price reference period, or treat it as a missing price and impute a price using the price movements of similar products. Where a CPI is revisable, it can be recompiled and the corrected figure published the following month. In some countries, the CPI is first published as a provisional figure to facilitate the late take-on of data including the situation just described.

5.90   Collectors should be encouraged to give feedback to the head office on their experiences of price collecting. Collectors are a valuable source of information and often give good early feedback on changes in the marketplace and can often warn of size or product changes before the head office is able to obtain this information from other sources such as newspaper advertisements. Collectors' feedback can be used to support observed price movements and to provide supplementary briefing material.

5.91   Feedback can also form the basis of a newsletter for collectors. Collectors' shared experiences can guide other collectors on how to treat different situations or circumstances.

5.92   The periodic routine of collecting prices in the field needs to be carefully planned and monitored, with arrangements in place to reflect local conditions. However, price collectors should send in information when it is due, and late submissions require follow-up.

Quality Checks of Local Price Collection: The Role of Auditors

5.93   One way of monitoring the work being carried out by price collectors and addressing any issues is to employ auditors to occasionally accompany collectors during the field collection or to carry out a retrospective check on the collected data. The function of the auditor is to check the validity of the prices collected and to initiate corrective action that may extend beyond correcting an individual price quote to reviewing and updating instructions to price collectors and to general retraining. The function can cover more than one geographical area, but it does not normally extend to the supervisor's role in managing price collectors and the price collection process. Sometimes the function of the auditor is combined with that of the field supervisor. The observations and comments of auditors are an essential part of quality management.

5.94   The range of tasks that an auditor carries out will vary from one NSO to another. Monitoring the standard of price collection will always be the main focus of the auditor. There are several other areas, however, in which auditors can be called upon to contribute. Auditors may be required to help with the sampling of locations and items and check that proposed collection locations contain an adequate range of outlets, and advise on economic conditions in these locations and on any dangerous areas. Auditors can carry out product reviews. For example, if an item is causing difficulty for price collectors, auditors can speak to collectors and retailers to determine the reasons for these difficulties. Auditors can also advise on changes to basket composition, ensure that products suggested by the head office are available across the country, and suggest item descriptions. Furthermore, auditors can provide reports on price collection in existing locations. For example, the head office may raise a query about an outlet in a location, and the auditors can visit this outlet to find the answer to the question or to persuade a retailer to continue with the survey.

5.95   The main purpose of audits is to ensure that each collector is following the procedures laid down for price collection so that the risk of errors is reduced. However, there are other benefits of strategic importance with regard to continuous quality improvement:

- Raising awareness of quality
- The identification of the scope for introducing improvements to quality, including rectifying weaknesses in procedures, documentation, and price collection skills

Quality Checks of Local Price Collection: Back-Checking

5.96   Another approach to monitoring the standard of price collection is to carry out a back-check (that is, a retrospective check of a proportion of the prices recorded during the collection).

5.97   Back-checks can be used to:

- Assess the standard of competence of individual price collectors
- Audit the overall standard of price collection
- Identify general training needs or the specific needs of an individual
- Highlight any key issues including, for example, problems with documentation or instructions issued by the head office
- Identify areas where collection is problematic; for example, all collectors may have problems in certain types of outlets, prompting the need for more detailed instructions from the head office

5.98   Back-checking should be done by an expert independent of the process (preferably employed by the NSO), such as an auditor. Back-checking is carried out by visiting the selected outlet and recollecting the prices and other relevant information, such as attribute or description codes. This activity should be carried out close to the original collection period to avoid problems of price changes occurring in the interim. Back-checkers should seek permission from the outlet staff beforehand and follow the general criteria of conduct for local collection.

5.99   Performance criteria should be determined to which all back-check results can be compared. These criteria

should set, for example, the acceptable number of pricing errors per number of items checked. Well-defined criteria will enable performance of collectors or locations following a back-check.

5.100   A back-check may include a range of tests to identify the following:

- Price difference—if the price is different, the auditor should check with outlet staff if there has been a price change since the original collection took place

- Insufficient item description—incomplete descriptions should be augmented to include all price-determining characteristics

- Wrong item priced—such as incorrect size or brand being chosen

- Items wrongly recorded as missing or temporarily out of stock

5.101   A report should be sent to the head office for scrutiny once the back-check has been completed. The head office will then need to take appropriate action, which may include, for example, retraining the price collector or sending out supplementary instruction.

Quality Checks Conducted Centrally by the Head Office

5.102   Four kinds of regular checks are necessary at the head office:

- Check that the price collectors' reports are sent in when they are due. If this is not done, it is necessary to find the reason and take appropriate action to obtain the reports.

- Confirm that the reports contain what they are supposed to contain (for example, mandatory fields have not been left blank, numeric fields contain numbers, or nonnumeric fields do not contain numbers).

- Review and edit each return. Substitutions may have to be made centrally, or those made by the collectors may have to be approved. Unusual or large price changes may need to be queried. Items priced in multiple units or varying weights may have to be converted to price per standard unit. Missing prices must be dealt with according to standard rules relating to the cause.

- Find and correct errors introduced when keying the numbers into the computer or transcribing them onto worksheets. Errors preferably should be avoided by eliminating the need to transcribe.

5.103   The way the data are organized in worksheets or in the computer may differ from the way they are arranged on receipt. For instance, the data may arrive at the head office organized for price collection purposes by collector, outlet, and item but will be entered on to a spreadsheet designed to reflect the computational needs of CPI compilation. The data in original format should be recorded for reference if any problems with the data are disclosed during processing. This facilitates operational management when dealing with queries. Furthermore, even if the same set of codes are used in price collection and in the processing of collected prices, other codes may have to be used for information that comes in from the collectors in noncoded form.

5.104   The organization of the quality checks conducted centrally by the head office will vary from country to country. In some cases, local or regional supervisors will do some of it; in other cases, it will be more appropriate for it all to be done centrally. Some of these tasks can be done by computer and others manually.

5.105   Procedures should be in place to check that all documents, messages, or files are returned from the field, so that price collectors can be contacted about missing returns. Initial checks should then be carried out to ensure that data are complete and correct. For instance, checks should be run to ensure that unexpected duplicate prices (that is, for the same item, in the same outlet, in the same location) are not taken on, and that the location, outlet, and item identifier codes, which accompany each price, exist and are valid. If any prices fail these checks, a query should be raised with the price collector for clarification. Since some of the checking may require reference back to the price collectors (or to their supervisors or respondents when direct mail questionnaires are used), the timetable for producing the index must allow for this communication to take place.

5.106   In deciding on what checks should take place in the head office, account should be taken of the validation checks carried out in the field. The use of handheld computers or tablets will increase the potential for validation at the time of price collection and reduce the need for detailed scrutiny at the head office. It would clearly not be productive or cost-effective to repeat all the tests already carried out locally, except as a secondary audit or random check that those checks have been completed.

Data Reports

5.107   Reports can help the head office staff identify prices for which the level or change stands out as different from that reported for similar varieties elsewhere, or simply where a change may need to be queried because it lies outside specified limits. Thus, a comparative analysis of other prices collected for an item can be undertaken. A printout can list all prices that either fall well outside the range of prices obtained for that representative item or for which the percentage change from last time falls outside a specified range, and a similar list can be compiled identifying outliers based on the recent price behavior of the same item in the same outlet. The limits used will vary from item to item and can be amended in the light of experience. The CPI compiler can then work through these lists, first ascertaining whether there has been a keying-in error, and then examining whether any explanation furnished by the collector adequately explains the divergent price behavior or whether a query should be sent back to the supervisor or collector. Again, the timetable for CPI compilation should allow for this, and anomalous observations should be discarded where an acceptable explanation or correction cannot be obtained in time.

5.108   Other reports may be produced regularly covering several periods (for example, three monthly) to detect accumulated patterns, thus enabling broader problems to be detected. For example:

- One collector's reports might show many more "outlet closed" remarks than those of other collectors, perhaps indicating either a motivational or training need on the part of that collector, or a change in retail trade patterns in a particular area.

92

- Variety substitution for a particular representative item might become more numerous than before, suggesting a possible need for revision of the specification or the choice of another representative item.
- Where tight specifications list several brands and models of which one is to be chosen, but a large number of collected prices are for items not specified in the original list, this suggests that the specified brands and models are no longer appropriate and that a review of the list is required.
- The dispersion of price changes for a specific representative item might be much larger than it used to be, raising the question of whether it has been appropriately specified.

5.109   Routine computer-generated reports should enable to detect such problems. Two types of reports are particularly useful: index dispersion reports and price quote reports:

- *Index dispersion report*. This is a list of the current index for each elementary aggregate, the number of valid quotes for each item, and the number of price relatives and their values. The ratios of current to previous valid prices can be compared and queries generated, if these ratios fall out of acceptable ranges based on previous price behaviors and considering any special circumstances such as the introduction of discounted prices in seasonal sales. The index dispersion reports can be used to identify quotes with price relatives that fall outside the range of the main bulk of quotes. These suspect quotes can then be investigated, and appropriate action is taken.
- *Price quote report*. This consists of a range of information on an item that the index dispersion report has highlighted as warranting further investigation. Information listed may include current price, recent previous prices, and base price, together with locations and types of outlet. The report can be used to identify the quotes that require further investigation and to investigate rejected prices.

5.110   While price collectors should examine every price they collect, it is not considered necessary nor feasible for collection supervisors and index compilers to subject every collected price to the same level of scrutiny. It is recommended that, to improve cost-effectiveness, a significance rating should be applied to determine how much time and effort should be expended on examining and, where necessary, editing individual prices.

- In general, prices from elementary aggregates with relatively small price samples should receive more attention from the index compiler. This is because, if the weights of the elementary aggregates are broadly equal, each individual price movement within these elementary aggregates can have a much more significant influence on the index calculation than any individual price movement from within an elementary aggregate with many price quotes.
- Price samples from elementary aggregates with high expenditure weights should also be examined critically as a higher expenditure weight will cause all price movements within the sample to have a greater influence on the CPI.
- The highest risk is associated with elementary aggregates with relatively large weights but few price quotes and

with complex index construction methods. This situation is associated with utilities and other services that account for relatively large expenditures and where there may be only one or a few suppliers and prices are based on complex tariffs.

5.111   Some of the techniques discussed in paragraphs 5.116–5.156 work best when applied to large quantities of data and have the advantage of being automated with the intervals identified for closer examination being generated by the prices data. Abnormal individual prices such as sale prices, or price movements such as sale recovery prices, may be excluded from manual and automated procedures for the detection of outliers, particularly in the calculation of upper and lower bounds, as they are not representative of the general trend in prices. Nevertheless, such prices should be checked for accuracy, for instance by reference to previous price history. Automated checking can be applied to seasonal sale prices and prices for seasonal items.

5.112   Automated checking essentially performs the same basic filtering purpose for the identification of outliers as the manual techniques described previously. It is sometimes referred to as statistical checking in contrast with the manual techniques which are sometimes referred to as non-statistical checking.

Data Validation and Editing

Data Validation and Editing by the Head Office: Automated (Statistical) Checking and the Use of Algorithms

5.113   Data validation methods identify possible errors and outliers for validation. Errors are incorrect prices, while outliers can be defined as price movements that are exceptionally large compared with most movements. The purpose of data validation is to validate and confirm prices flagged as errors or outliers. Any errors should be corrected. Outliers verified as correct should be used in the calculation of the index.

5.114   The main conceptual difference between automated (statistical) checking and manual checking is that the automated technique calculates the limits for acceptable movement based on the data collected. These techniques have the benefit of automatically updating the acceptable limits in line with any overall change in price volatility, as observed when new price data are received and the limits are recalculated. These techniques require a large amount of data to provide reliable results and so are best suited to data handlers and index compilers in regional offices and the head office, where prices data from several collection centers will have been collated and stored, rather than at the local level.

5.115   Automated (statistical) checking compares each price change with changes in the other items from a given price sample. The chosen price sample is usually the sample to which the item being checked belongs, but the sample for testing may be a combination of price samples for similar products. It can also be updated as more prices are received from the field. For each of the methods described in the following text, the price ratios may measure the price change over any time period: for instance, the change from the

previous period or the change from the same period in the previous year.

The Use of Median and Quartile Values

5.116 One method of setting the limits to determine whether a movement is a possible error is based on the median and quartile values of the price ratios (R) from the sample. The acceptable limits are set as a predefined multiple of the range between the median and the quartiles. Any observation with a price change outside this range is identified as a possible error. The major benefit of a method like this is that it is not affected by any single outlier value. A numerical example is provided in Annex 5.3.

5.117 The basic approach to estimating sensible upper and lower limits of acceptable price movement relies on the assumption that the observed price changes are normally distributed. Under this assumption, the distance between each of the first and third quartiles (R_{Q1} and R_{Q3}) and the median (R_M) will be the same: call this distance "D_M." Operating under this assumption, the proportion of price changes that are likely to lie outside specified upper (L_U) and lower (L_L) limits can be estimated from a normal distribution table. The limits can be defined as

$$L_U = R_M + C \times D_M; \text{ and} \qquad (5.1)$$
$$L_L = R_M - C \times D_M$$

where C is a user-defined value.

5.118 As discussed in paragraphs 5.126–5.134, a variation of this basic approach is recommended to allow for the skewed distribution of price changes that can be observed in practice.

5.119 If C is defined as equal to one, then approximately 50 percent of the observations will lie between the upper and lower limits. Using the standardized normal distribution, this is equivalent to setting the limits at plus or minus 0.7 times the standard deviation (σ) from the median. Table 5.2 provides approximate multiples of σ for selected values of C and the associated percentage of the observations that will be flagged as possible errors and outliers. In practice, there are serious shortcomings with this method as described here.

5.120 In normal circumstances, the majority of observations for many products will not show any price movement. Therefore, the values of the quartiles are likely to be very close to the median value. As a result, using small values for C is likely to cause the majority of price movements to be flagged as possible errors and outliers. To demonstrate this effect, in Example 2 in Annex 5.3, 16 additional observations indicating no price movements have been added to

the sample of 30 observations used in Example 1. A price sample with at least one-third of the observations showing no movements would not be unusual for many categories of items. If C is set to two, then 60 percent of the actual price movements would be flagged as possible errors, compared with 30 percent in the unadjusted sample.

5.121 The index compiler should experiment with different values of C for different product groups or outlet types to determine appropriate values for local use. It is recommended that a relatively low value of C be used. C need not be an integer and can be expressed with regard to fractions as well.

5.122 The distribution of prices and price movements is rarely normal; rather in most cases, a skewed distribution exists. Thus, the underlying assumption of a normal distribution is not valid, and the use of symmetrical upper and lower limits will result in a skewed distribution of prices flagged up as possible errors or outliers. This is operationally inefficient and the examination of differing proportions of "low" and of "high" prices and price movements could lead to bias.

A Modified Use of Median and Quartile Values

5.123 To use the previous method in practice, three modifications are recommended, as shown in paragraphs 5.127–5.134.

5.124 Based on the simple price ratios, the distances from the median represented by price decreases are not as large as the distances represented by price increases. As an example, consider a case where a product is on special offer at half price. This is represented by a price decrease of 50 percent. However, to return to the original price requires a 100 percent increase. To make the calculation of the distance from the center the same for extreme changes for price decreases and for price increases, the price ratios should be transformed. The transformed distance, S_i, for the ith price observation can be calculated as

$$S_i = 1 - \frac{R_M}{R_i} \qquad \text{if } 0 < R_i < R_M \qquad (5.2)$$
$$S_i = \frac{R_i}{R_M} - 1 \qquad \text{if } R_i \geq R_M$$

where

$$R_M = \text{median.}$$

5.125 The observations with a price ratio lower than R_M have now been transformed into the negative of the increase required to return the price ratio to the value of R_M. Any observations with a price ratio equal to R_M will have a transformed price movement of zero. Observations with a price ratio greater than R_M have been transformed to show changes as though they had increased from R_M. The procedure is then carried out on the set of S_i.

5.126 In situations where the quartiles (R_{Q1} and R_{Q3}) are quite close in value to the median (R_M), many small price movements are likely to be identified as possible errors or outliers. To reduce this problem, items with no price movements may be removed from the calculations. If the acceptance interval is still very narrow, some minimum distance

Table 5.2   Selected Values of *C* and the Proportion of Observations Flagged

C	σ Multiplier	Expected Proportion of Observations Flagged (percent)
1	0.68	50.00
2	1.37	17.00
3	2.07	4.00
4	2.75	0.70
6	4.00	0.14

94

should be set. A starting value is 5 percent for monthly changes but it is up to the CPI compiler to choose, based on past experience.

5.127 The third modification is intended to overcome the problem of using small samples. When using a small sample, the impact of one observation on the distances between the quartiles and the median might be considered too significant. In practice, the sample sizes for many elementary aggregates will be small. To improve the usefulness of this method, the samples from several similar elementary aggregates can be combined. In this regard, elementary aggregates can be considered similar if their prices are believed to exhibit similar behavior.

5.128 The Hidiroglou and Berthelot method, as described in detail in Hidiroglou and Berthelot[6] (1986), can easily be extended according to the description in paragraphs 5.133–5.135. The variable s_i is independent of the price levels. To address the issue that the level of the prices can influence the acceptance interval, s may be transformed into a new variable, E:

$$E_i = s_i \cdot \left(\max \left\{ p_i^{t-1}, p_i^t \right\} \right)^U, \, 0 \le U \le 1 \qquad (5.3)$$

5.129 E is calculated as s multiplied by the largest of the prices in period t or $t-1$, raised to the power U. The variable U determines to what degree the price level influences the acceptance interval. The larger U is, the larger the influence of the price level will be. If $U = 0$ the price level plays no role. This method is helpful, if the compiler wishes to pay more attention to a price increase from 1,000 to 1,100 than from 10 to 11.

5.130 A further transformation can be made to ensure a minimum acceptance interval to avoid that too many price changes being identified as possible errors for elementary aggregates (E) with only small price changes. In this case, for each elementary aggregate, the median, E_M, and the first and third quartiles, E_{Q1} and E_{Q3}, of the E_i's are found, and the following values calculated:

$$d_{Q1} = \text{Max} \, \{ E_M - E_{Q1}, |AE_M| \} \qquad (5.4)$$
$$d_{Q3} = \text{Max} \, \{ E_{Q3} - E_M, |AE_M| \}$$

5.131 A is a constant that enters $|AE_M|$ to ensure a minimum acceptance interval. A low value of A raises the probability that $(E_M - E_{Q1})$ or $(E_{Q3} - E_M)$ determines d_{Q1} or d_{Q3}, and vice versa. For instance, if A is set to 0.05, $|AE_M|$ will be quite small so that $(E_M - E_{Q1})$ or $(E_{Q3} - E_M)$ are likely to determine d_{Q1} or d_{Q3}, even if the dispersion of the E_i's is relatively small. If, on the other hand, the dispersion of the E_i's becomes very small, $|AE_M|$ determines d_{Q1} and d_{Q3}. Hence, A can be used to avoid having too many price changes identified as possible errors for elementary aggregates with only small price changes. The acceptance interval is finally defined as

$$\text{Acceptance interval} = \{ E_M - C^* d_{Q1}; E_M + C^* d_{Q3} \}, \quad (5.5)$$

---

[6]Mike Hidiroglou and Jean-Marie Berthelot. 1986. "Statistical Editing and Imputation for Periodic Business Surveys." *Survey Methodology* 12 (1): 73–83.

where $(E_M - C^* d_{Q1})$ is the lower bound, and $(E_M + C^* d_{Q3})$ is the upper bound of the interval. C is an extra variable that may be introduced. The larger the C, the larger the acceptance interval, and the fewer extremes and potential errors will be identified.

The Tukey Algorithm

5.132 The Tukey algorithm overcomes the problem of validating data when there are many observations with no price change (that is, where many price relatives are equal to one indicating no price movement). The first step is to sort the sample of price relatives. The highest and lowest 5 percent are flagged for examination as possible errors or outliers and removed from further calculation. All observations with no price movement are also removed from the sample before further calculations are done. The next step is to calculate the arithmetic mean (AM) of the remaining observations (referred to as the Tukey sample). This value is then used as the dividing value to separate the observations into two smaller samples: an upper and a lower set of price ratios. The arithmetic mean of each of these two samples is then calculated as (AM_L, AM_U). The upper and lower Tukey limits (T_U, T_L) are then calculated for the Tukey set as

$$T_U = AM + 2.5(AM_U - AM) \qquad (5.6)$$
$$T_L = AM - 2.5(AM - AM_L)$$

5.133 All observations that are greater than T_U or less than T_L are flagged as possible errors or outliers.

5.134 As this method excludes all observations with no price movement, the calculated limits are unlikely to be close to the mean. Therefore, there will be no need to impose a minimum difference. However, the problem of requiring a reasonably large number of observations in the sample remains. Again, it may be necessary to combine the samples of similar elementary aggregates. Example 3 in Annex 5.3 shows that five observations would have been flagged by this method in comparison to 18 observations by the previous method based on the modified use of median and quartile values (see 5.126–5.134).

5.135 As mentioned before, statistical methods of filtering have an advantage over simple filtering, because the limits are set by the data and can be recalculated over time. The disadvantage is that filtering cannot be done until sufficient quantities of data have been collected, unless the index compiler uses approximations from past experience. The processes can be repeated as additional prices are received. Compilers should aim to set filters so that most of the records flagged as potential errors do turn out to be errors (or outliers requiring explanation). The aim of all these methods of filtering is to indicate which records require examination, not to flag records for automatic deletion from the sample. Each price movement should be checked for credibility and representativeness. Only if the movement is an error or unrepresentative should modification be considered. There should not be a presumption that an outlier is "wrong until proven right," and outliers should not be treated as incorrect prices.

95

Visual Data Validation

5.136 Using plot charts is helpful to spot the outliers on the collected data and focus validation on these (see Figure 5.1 with arrows highlighting outliers). For visual data validation, it is easy to use plot charts that are readily available in spreadsheet software tools or can be programmed on the information technology (IT) system. Visualization can be more convenient than focusing validation to the change of prices, especially when there are extreme price changes such as during sales periods or in the case of fresh fruit and vegetables. If the plot chart is programmed into the IT system, from the outlier it is possible to make a direct link to the observation.

Review of Outliers

5.137 The detection of price observations that are outliers may be conducted through an examination of both price levels and of price movements. The movements will have been verified as being based on correctly collected and recorded data but may not be representative of the behavior of the section of the market that they are meant to represent. This leads to the concern that a different sample would have produced a significantly different and more representative average price movement.

5.138 It is recommended to use resources efficiently on validation and checking of input data and to focus on identifying the most important errors/outliers. The general rule should be to include verified prices. Excluding or modifying prices should be the exception. The aim should be to reflect the reality.

5.139 The tests for outliers are the same as those for identifying potential errors. Outliers can be determined by comparing the price movement relative to defined allowable limits. These may be either predetermined numerically or predefined based on statistical tests.

5.140 If, by exception, outliers are to be modified, they are usually modified to lie on the predefined boundaries of acceptable movement or to be imputed by the movement of a suitable sample of prices. Imputation by the average price change of the product group to which the item belongs yields a similar result as its exclusion (the same result within the elementary aggregate), but such imputations can have operational advantages as they employ protocols already in the calculation system for the imputation of missing prices. An automatic adjustment should generally be avoided, and not be used to reduce volatility in an index, for example, at the elementary aggregate level. The index compiler should consider each case on its individual merits, following agreed guidelines and deciding based on all relevant information. Prices should be modified or discarded only if there is sufficient justification. The CPI protocols followed by the NSO may even forbid the modification or exclusion of outliers.

5.141 Price collectors and their supervisors are responsible for providing as much information as possible about the reasons for extreme price movements or levels and why they accepted the price quote as valid. In addition to checking for better accuracy, supervisors can also be instructed to compare the price movements for equivalent products obtained by all the collectors they supervise.

Figure 5.1   Price Changes in Plot Chart during Sales Season
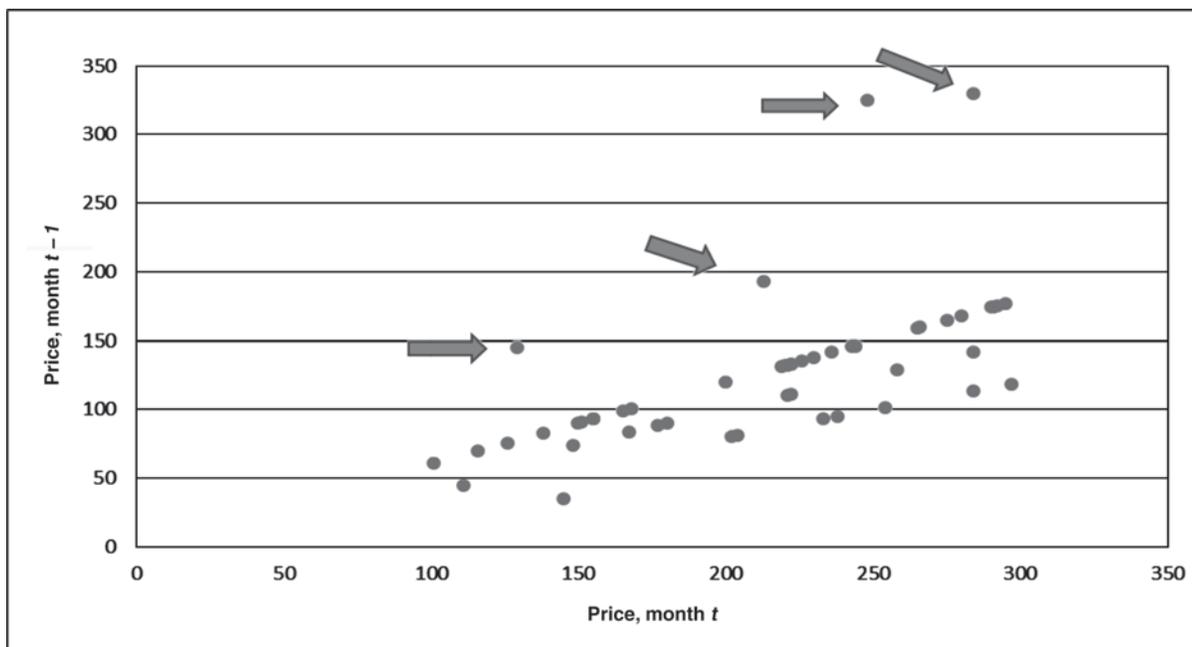
No Price Change

5.142   If the price observations are collected in a way that prompts the respondent with the previously reported price, the respondent may report the same price as a matter of convenience. This can happen even though the price may have changed, or even when the particular product being surveyed is no longer available or has changed its price-determining characteristics. As many item prices do not change frequently, this kind of error is unlikely to be spotted by normal checks. Often the situation comes to light when the contact at the responding outlet changes and the new contact has difficulty in finding something that corresponds to the price previously reported. It is advisable, therefore, to keep a record of the last time a respondent reported a price change.

Missing Prices

5.143   Treatment of missing prices is dealt with in more detail in Chapter 6. This section discusses ways of minimizing the occurrence of missing observations.

5.144   It is important to maintain the relevance of the sample of items priced. As part of the longer-term maintenance of price samples, items and locations for which prices are missing can be examined for common patterns. For instance, if many retailers are missing the same item, there may be a general supply problem. This may be an indication that an item will have to be replaced. If the number of regularly missing items is growing, then the sample might need to be reviewed. If a particular outlet is recorded as having a relatively large number of missing prices it may no longer be appropriate for the particular items assigned to it, or the varieties whose prices are collected in the outlet may need to be reviewed.

5.145   When prices are collected using a questionnaire sent to the outlet, individual respondents generally follow a regular pattern in terms of response. Some will return their price survey promptly, while others will take more time. Price collectors should be encouraged to become familiar with these patterns. If the system for recording the return of these surveys also records the expected return date, then unexpected nonreturns can be flagged even though the final deadline for return of survey forms has not passed. These respondents can be contacted in advance of the final deadline to ensure that the survey form has not been forgotten. Early contact can reduce the number of prices still missing by the deadline. Respondents that provide prices for heavily weighted items can also be monitored and contacted earlier.

Credibility Checking

5.146   Credibility checking tests the reasonableness of the input data and the results obtained. Credibility checking of the results should take place after the checking of the numerical accuracy of the data at or shortly after price collection as described previously. These early checks are the responsibility of the price collectors and their supervisors but also involve outlier detection at the head office. These early checks should discover all straightforward errors like incorrect coding (for example, wrongly attributing a price as a sale price) and the incorrect recording of prices.

5.147   Addressing other potential errors is less straightforward. Results that fail a data check, such as exceeding the predefined movement limits described earlier, may be judged by the index compiler to be valid as a result of referring to other information such as market intelligence. Other potential errors might only be resolved after checking with the respondent, if time allows.

5.148   If it is possible with individual price quotations to resurvey the price or obtain a satisfactory explanation from the respondent, the query can be sent back to the price collector and the data can be flagged as being verified and then subsequently corrected if found to be in error. Even if it is not possible to check with the respondent before the computation deadline, the respondent could be questioned during the next regular visit, as the answer may assist the NSO's understanding of market behavior for the particular product or retail sector. When a satisfactory explanation is not available, the CPI procedures should provide guidelines to aid the compiler in deciding how to treat the questionable price. For instance, the compiler could omit the price, allowing the processing system to impute a price, or modify the price to keep the price change within a predefined limit, but this is best avoided and should be the exception. If prices are modified without verification from the respondent, it is recommended that price collectors be informed of potential problems during the next collection.

5.149   NSOs can minimize problems caused by unusual prices and price movements by training price collectors to recognize these situations, to check prices when first observed, and to collect relevant explanatory information during the initial price collecting visit. Avoiding return visits or calls keeps costs down and reduces the burden placed on respondents.

Checking by Impact or Data Output Checking

5.150   Filtering by impact, or output editing, 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. The impact that a price change has on an index is its percentage change times its effective weight. However, the exact calculation of the impact will depend on which formula has been applied for the elementary indices. It is possible to set a maximum value for this impact, so that all price changes that cause an impact greater than this can be flagged for review. The impact of a price change on a higher-level index will also depend on the weight of the elementary index in the aggregate.

5.151   At the lowest level, the appearance and disappearance of products in the sample cause the *effective weight* of an individual price to change substantially. The effective weight is also affected if a price observation is used as an imputation for other missing observations. The evaluation of effective weights in each period is possible, though complicated. To help highlight potential errors, nominal weights, as a percentage of their sum, will usually provide a reasonable approximation. If the impact of 12-month changes is required to highlight potential errors, approximations are the only feasible filters, as the effective weights will vary over the period.

5.152   One advantage of identifying potential errors in this way is that it focuses on the results. Another advantage

97

is that this form of filtering helps the CPI compiler describe the contributions to change in the price indices. Much of this analysis is done after the indices have been calculated, as the CPI staff often wishes to highlight in the statistical press release those indices that contributed most to overall index changes. Sometimes the CPI staff findings that some retail sectors have a relatively high contribution to the overall price change may be considered counterintuitive. The change may also be traced back to an error, but it may be late in the production cycle and jeopardize the scheduled release date. There is thus a case for identifying such unusual contributions early as part of the data editing procedures rather than for analytical purposes. The disadvantage of this method is that in practice the final calculation of an elementary index change may be rejected only after the CPI has been computed.

Price Collector Training

5.153   The training of local price collectors and clear instructions for them are vital elements in ensuring the quality of the prices data and of the CPI. Collectors need to be properly trained, require adequate instructions, and must have easy access to guidance because:

- Price collection is of significant policy relevance.
- Quick judgments often need to be made.
- Collectors often work remotely and on their own.
- Instant communication is not always possible.
- Collectors work in a dynamic environment.
- Errors are difficult to rectify.

5.154   Documents are needed to explain what is to be done, when it should be done, how it should be done, and why it should be done. Reviewing the documentation also provides an opportunity to review the procedures.

5.155   Good documentation as part of an integrated quality management system is addressed in Chapter 13. The current chapter deals specifically with the documentation needs of price collectors and training.

5.156   Training for price collectors should enable them to successfully perform all essential activities and deal with potential difficulties including:

- Persuade new outlets to become price providers
- Understand and recognize occasions when prices provided are unacceptable
- Record relevant information to describe the quality change in a product
- Recognize unusual price movements when checking their collected prices

Training for Price Collectors

5.157   Introductory training should be given to all price collectors so that they gain the necessary skills before collecting prices for the CPI. It can also be a motivational tool.

5.158   A typical training schedule might consist of a one-day training course at the head office (which may include some refresher training for experienced collectors) covering:

- Background to the NSO and the CPI
- Use of the CPI and importance of accurately recording prices
- The general principles of index compilation and price collection
- How local price collection fits into the overall CPI compilation process
- Instructions for retailer recruitment, getting permission to enter outlets, and so on
- *Practical price collection issues*: for example, product identification or descriptions; pricing (for example, item descriptions, definition of price/sale price, rules relating to seasonal items, quantity conversions, quality adjustment—when is an item or product equivalent?)
- The timetable and administrative arrangements

5.159   Practical examples and practice collections should be an integral part of the learning process. For example, there should be an opportunity for:

- Discussions about "equivalent" replacements using photographs and item descriptions
- Practice collections in office
- Supervised practice collection in the field

5.160   Tests and evaluation of individual performance should be an integral part of the training. This could be achieved through:

- Written tests at the end of the training day
- Evaluation by supervisors of practice collections in the field
- Feedback to new collectors, including additional training needs
- Evaluation by collectors of training provided (essential for ensuring training is relevant and effective)

5.161   The evaluation of individual collector performance is essential. Collectors must pass the required standards, against a checklist of tasks, before being allowed to conduct a real collection.

Follow-Up Training and Refresher Training

5.162   The longer-term training is just as important as introductory training to the integrity of price collection, particularly with the evolution of the retail sector and CPI methodology and when CPI baskets are updated. One way of facilitating this is for the price collector's supervisor to:

- Accompany the new price collector on a live price collection.
- Conduct a back-check of the prices collected to identify any problems.
- Produce an evaluation report that will provide a basis for further training of the price collector. The evaluation report can include a scorecard against a checklist of required actions.

5.163   Where resources allow, NSOs should conduct regular accompanied checks and background checks of all price collectors in addition to special checks of performance issues. The information gathered can be used to compile

98

scorecards for individual price collectors, supervisors, and groups of price collectors.

5.164 Regular refresher training workshops should be considered, especially where evidence from the field indicates a need or where price collection procedures and conventions change, or the CPI basket or sample of outlets has been updated. These present an opportunity to: raise awareness of the importance of collecting correct prices; provide formal training on revised guidelines; resolve recurring or recent problems; and provide price collectors with an opportunity to assist each other in managing problem situations encountered in the field, such as dealing with reluctant respondents.

Training of Supervisors and the Head Office Staff

5.165 Supervisors must be at least as well informed as the price collectors. As the supervisor will normally be the first point of contact when a difficult situation is encountered during price collection, they also need a good understanding of the methodology and theory behind the selection of the product sample. Supervisors are part of the management team. Their training should cover:

- Team management
- Performance appraisal (where this is normal office practice)
- Project management

5.166 The head office staff should also be provided with a basic training in price collection. The benefits are threefold:

- It gives the head office staff a better understanding of collectors' needs.
- It helps editing (staff at the head office will know what to look out for).
- It supports disaster recovery (paragraphs 5.175–5.177) or business continuity (the head office staff will be able to undertake price collection in an emergency).

Documentation: Work Instructions

5.167 Accessible, relevant, and up-to-date work instructions are essential for both price collectors and their supervisors. Documentation should cover all aspects of the job and should in large part reflect what has been covered in training. Price collectors should be provided with work instructions on:

- How to approach outlet staff
- How to ask questions to ensure that the required information is obtained
- Appropriate personal behavior and dress codes
- Procedures for recording and passing on collected prices and other relevant information
- Data checking
- Creating collection schedules
- Recognizing when recorded prices appear to be incorrect

5.168 The work instructions for supervisors of price collectors should be in the form of a supplement to the price collectors' work instructions and should cover:

- Checking the quality of the price collectors' work
- Checking the accuracy and completeness of the prices collected
- Official recording of resource use (for example, cars and bicycles for transport and funds for buying goods in markets)
- Official procedures for maintenance of resources (for example, testing the accuracy of scales)
- Creating complementary collection timetables for all collectors within the supervisor's area of responsibility

5.169 Most of the documentation should be prepared by the head office, with input as appropriate from regional offices, fieldwork supervisors, and data collectors. Centrally prepared documentation will help ensure consistent practices in the field including between regions and should be readily available to all collectors and supervisors. The documents can be available in paper or electronic form and should be accessible to the relevant staff.

5.170 All documentation should be kept up to date. Effective documentation control systems should be in place. With paper-based documentation, this could mean keeping the instructions in a loose-leaf folder and issuing individual updates. The amendment pages should include version number and date printed and be kept to a reasonable number for ease of reference. Editorial access should be restricted and password protected. A judgment will need to be made on when a redraft of a chapter or the complete working instructions is justified. Documentation is an essential part of a quality management system.

5.171 An example of a documentation control template is given in Annex 5.4.

Disaster Recovery

5.172 The prices data should be stored on a database with hardware and software that is robust and supported to minimize the business continuity risks associated with running the existing system. But even with a resilient production system, contingency planning and operational continuity in the field and the head office are essential.

5.173 In a world of rapidly changing statistical needs, a statistical system should be able to respond quickly and effectively to changing demands and should have the resilience needed to ensure continuity in the production of statistics. This is not possible if systems are old, inflexible, and extensively tailored to past requirements. Building a modern statistical infrastructure of methods, tools to implement them, and a technical environment to support the statistical processes is a significant component of achieving quality CPI compilation and computation. Contingency plans are needed when the unexpected happens, for example, when there is a systems failure or the price collection team is affected by an unexpected illness. Disaster recovery plans address these risks by, for example, saving regular and frequent backup copies of the prices database on a secondary computer and by having the capacity to collect prices when significant numbers of

99

price collectors are not available. Two strategies have been followed to accommodate a short-term unexpected deficit in the numbers of price collectors:

- The collection of prices from a subsample of outlets chosen to be representative of the full sample, thereby not needing so many price collectors. The price evolution from one period to the subsequent period can then be calculated from the subsample using matched pairs of price observations.
- The training of the head office staff in price collection (for example, as part of their initial training), so that they can provide cover. The head office can also be given responsibility for price collection on a routine basis in a location close to the head office, with individual head office staff allocated the task of price collection on a rotating basis. An added benefit of this is that the head office staff become more familiar with the issues confronted by price collectors.

5.174   In cases where no fieldwork can be performed, for example, where there is an all-out strike and no prices can be collected by visiting outlets, an indicative figure may be possible using data from other sources, such as retailers websites, but only if it can be established that like-for-like comparisons are being made and that the fixed-basket principle is being adhered to.

Other Methods of Price Collection

5.175   This chapter has so far focused in large part on traditional methods of price collection where price collectors visit outlets and record prices on paper forms. It now considers other methods of price collection. As noted at the beginning of the chapter, scanner data are given a separate consideration in Chapter 10.

Electronic Reporting

5.176   Electronic reporting for centrally collected prices and use of handheld computers and tablets for local price collection can introduce greater efficiency into price collection and processing, and provide more scope for effective quality assurance of prices and auditing, but both depend on the introduction of effective quality control procedures. Electronic reporting through electronic point of sale, commonly referred to as scanner data, is another option.

5.177   Centrally collected data can be collected electronically in several ways. Once initial contact has been made with data suppliers, a mutually convenient electronic data collecting procedure can be initiated. Options include:

- Emailing data collection spreadsheets between the NSO and the retailer
- Emailing of price lists at agreed times by retailers
- Touch-tone dialing facilities for data to be supplied in an agreed-upon format
- The use of the internet (supplemented, if necessary, by telephone calls to clarify definitions and availability and whether the prices displayed on the internet are the same as those displayed in the corresponding outlets) (see also paragraphs 5.192–5.206 on collecting prices online and web scraping and Chapter 10 on scanner data)

5.178   The use of electronic websites/portals, where respondents can report prices online, is becoming increasingly

common and can also be an excellent channel for communicating with the respondents.

Collection by Telephone

5.179   The prices for certain items, particularly services such as electricians' and plumbers' charges and the cost of home security, may be obtained by telephoning the business or organization concerned. This applies when the outlets provide standard items or services. However, even if prices are obtained by telephone, the outlet should be visited occasionally. This helps to maintain cooperation through personal contact and to ensure that there are no misunderstandings over the prices. This will be more important for some outlets than others. For example, the price of hiring a van may be less certain than the cost of an eye exam.

Computer-Assisted Data Collection (CADC): The Use of Mobile Telephones, Handheld Computers, and Tablets

5.180   A number of NSOs have successfully used mobile telephones, handheld computers, or tablets for local price collection. These technologies are now available at competitive prices and the necessary infrastructures are generally in place to make the use of CADC an attractive option.

5.181   A CADC system can lead to improvements in the quality of CPI data, particularly as increased quality control at the point of data entry helps identify anomalies and ensure that prices are correct. CADC has the potential to significantly improve the quality of the final CPI in the following ways:

- *Price history*. The price collection program might allow for a more comprehensive price history to be available to the price collector, rather than just one previous price included on paper forms. The availability of such data leads to less judgmental editing at the point of data collection and helps ensure the comparability of items, particularly where prices for a particular item are variable. Some price statisticians have argued that machines should be programmed to reveal the price history only after a price quote has been entered so that collectors are not overly influenced when locating the item to be priced or when choosing a replacement item. Others argue that an early sight of the price history is useful information for the price collectors as it assists them in their work.
- *Quality checks in the field*. The price collection program can include several automatic validity checks that can be used to identify where the price entered varies by a certain percentage (positive or negative) from the previous month's price and the average price for that item over a number of months and to flag up where data were not entered in all required fields (price, weight, indicator code). These checks provide a useful marker for when a price needs to be double-checked. In a paper-based system, such checks are carried out in the head office after the data have been collected, and audits can be carried out after the collection period when prices may have changed.
- *Transcription*. There is a major risk of errors when transcribing paper forms. This is not a risk when using CADC, where data can be transferred electronically to the head office.

5.182 The use of CADC also significantly reduces the time taken to make data available electronically at the head office and between data collection and finalization. This can be achieved through:

- *Transcription.* Data collected on paper must be transcribed onto a computer for computation. This process is time-consuming and resource-intensive. When data are collected on handheld computers or tablets, the data can be directly transferred electronically to the servers at the head office potentially in real time.

- *Transmission from regions.* Electronic transmission will allow price collectors or regional offices to directly transmit an electronic data file to the head office thus avoiding the need for postal or courier services or hand delivery forms. This significantly increases the speed of data transmission to the head office and reduces the cost of doing so. In addition, the head offices can look up the latest returns of price data from all regions immediately on receipt and identify early any issues.

- *Quality checks in advance.* As the functionality is available to run certain quality checks in the field that would normally be run in the office after data were transcribed, the time taken for quality checking centrally can be reduced, or, alternatively, extra supplementary checks can be carried out.

5.183 These improvements to the speed of the processing system can facilitate an earlier publication or provide opportunities to spend more time on analysis and interpretation, the production of press releases and associated briefing, or the collection of more prices.

5.184 A CADC system enables certain checks that improve the efficiency of the CPI management. These include:

- *A check that all prices have been collected before the collector leaves the outlet.* An electronic data collection form can easily check whether all prices have been collected and flag when they have not. This mitigates the risk of the price collector inadvertently forgetting to price an item.

- *A check on when prices were imputed.* Electronic data collection can automatically record a date/time when the prices were entered in the machine. This is useful for validation purposes.

- *Indicator codes.* CADC provides the opportunity for additional features to be included in the data collection form. One such feature would be indicator codes (represented by a single letter) that can be used to show when a price collected is for an item on sale, a replacement item, a missing item, a discontinued item, etc. This is a simple tool to enhance the ease of validation and the management of the item list (see paragraph 5.73).

Having price histories more readily at hand can:

- Make briefing of price collectors prior to fieldwork more effective, for example, by a better appreciation of when an "outlier" is a legitimate price change and vice versa.

- Add to the quality assurance processes through assisting with analysis when the index has been compiled and the briefing is being put together. These advantages are particularly relevant when there can be significant regional variations in price levels and trends.

5.185 There is a short-term cost associated with the introduction and implementation of CADC for CPI price collection. Costs include:

- The purchase of equipment.
- Upgrading "back-office" systems to enable interaction with the handheld computers, mobile telephones, or tablets.
- The development of appropriate software for local price collection building on the experience of others. Costs depend on the functionality and sophistication of the program. Some NSOs have developed software for CADC that they may be ready to share.
- Training of field staff and NSO staff on using the new systems, including pilot price collection.

5.186 There will also be longer-term costs associated with maintenance of the system and training of new staff, but the additional expenditure on the latter is likely to not be significant as new staff must be trained whatever system of price collection is used.

5.187 When planning to use handheld computers, mobile telephones, or tablets for local price collection, the decision needs to be made on whether the software is designed specifically for a certain hardware or not. If it is hardware dependent, the life cycle of the software is usually dictated by the life cycle of the hardware. In data transmission, the issue of confidentiality needs to be addressed as well as how to secure the transmission action in practice with regard to reliability. The speed and reliability of any network that is used should be tested from all price collection locations.

5.188 There are several advantages when using CADC, but as with all data collection methods it has risks and limitations. With CADC the first risks are with the devices: the battery might not last the whole day of price collection, especially if price collection is done in extreme conditions, and the data might be lost if the device breaks down during price collection period. Also, the devices have life cycles and if the software is coded to operate on a certain device this may need to be rewritten for a replacement device. This limitation causes costs when the devices come to the end of their life cycle. There are also risks in data transfer, especially if the connections are poor, for example, if for any unpredicted situation the connection fails during downloading the data from CADC to the database. Some of these risks can be avoided using different solutions, like programming the software independently from the device choice.

Collecting Prices Online and Web Scraping

5.189 A distinction should be made between collecting the prices of goods and services purchased online, to reflect the increasing importance of the internet as a channel for making purchases. A strategy to collect prices online through automation has the potential to reduce costs when compared to the resource-demanding process of manual price collection. Prices for online sales can differ from the prices charged at physical outlets—even for the same retailer—and the profile of goods and services purchased online can be different to other purchases. The prices in the

101

CPI must be representative and accurate. The motivation in adopting different data sources and extraction techniques is important. The motivation for collecting prices online can be twofold—efficiency and to ensure online purchases are properly representative.

5.190 Retailers with an online presence—either as sellers of goods and services online, or as retailers who do not sell online but use the internet to list prices—should be treated like any other retailer and be contacted first by the head office and be invited to participate in the price survey even though this will not involve a physical interface.

Online Collection of Prices

5.191 This section relates to collecting prices online from publicly accessible websites, referring to goods and services also sold in the corresponding physical outlets. It is a way of increasing the efficiency of price collection for a traditional sample of retail outlets and a fixed basket of goods and services. It does not relate to web scraping, the collection of prices for online purchases or the use of scanner data.

5.192 Instead of the traditional way of sending a price collector to a retail outlet, extracting prices online directly from websites will significantly reduce the price collection costs. Similarly, the response burden on data providers will be reduced to close to zero when extracting prices online directly from websites to replace postal or online questionnaires. Collecting prices online is relatively straightforward although care needs to be taken and checks put in place, especially when automated technical solutions are adopted. The prices obtained online must represent the transaction price in the physical outlet. Checks need to be made that the price advertised online is the same as the price advertised on the retailer's website and in the physical outlet and that no overhead associated with buying online, such as delivery charges, are included.[7] Also, online data need to contain sufficient information on characteristics to detect changes in quality. When using online data sources, checks also need to be made that product code numbers, if used to identify the good or service, have not changed between price collection periods and that the codes are unique.

5.193 Collecting prices online, which can be relatively easy and cheap, will not always be a suitable substitute for price collectors.

Collection of Prices for Online Purchases

5.194 Goods and services purchased on the internet need to be properly reflected in the sample of prices used to compile the CPI.

5.195 Elementary product groups should be stratified to reflect products purchased online. Sale information for weighting and the drawing of samples of internet purchases can be taken from HBSs, which should record information about outlet type (including internet purchases), and from information supplied by online retailers and market research companies.

5.196 The sample of items to be priced should be representative of all online purchases and will be different from online collection of prices charged by physical outlets. The prices recorded should represent the full cost of purchase. Online purchases may include standard extras such as delivery charges. Unavoidable charges that are directly connected to the purchase of the priced product and which are not separately invoiced should be included in the price for the purpose of CPI compilation. If the charge is separately invoiced or relates to the purchase of a number of items, then the treatment is less clear-cut. One option is to include these charges under transport services, but issues relating to the Classification of Individual Consumption According to Purpose (COICOP) arise (for more information on COICOP, see Chapter 2). Another option is to follow the approach used for the Harmonised Index of Consumer Prices in European Union countries, where such unavoidable charges that are not part of the basic advertised price may be considered as an inseparable bundle of a good and a service and can be treated as one product (for additional information, see the section on internet purchases in Chapter 11).

Web Scraping

5.197 Web scraping is the process of automated collection of data from the internet through a set of computer software techniques for extracting information from websites (webpages) or using an application programming interface, which is a set of routines, protocols, and tools for building software applications. Web scraping identifies and retrieves relevant data and downloads and organizes them in a suitable format for computing a CPI.

5.198 There are technical measures applied in some websites to avoid web scraping activities. These measures block the scraper Internet Protocol address access to the website or block the response to the scraper http browser agent identification. The action is triggered after identifying an "abnormal" behavior (by analyzing the activity log) or by filtering access from some agents (through the robots.txt server configuration file).[8]

5.199 Anti–web scraping devices by retailers reinforce the need to gain the cooperation of online retailers prior to data scraping to avoid being blocked from data collection. Retailers should always be informed about the nature, extension, and frequency of web scraping actions prior to any web scraping taking place. NSOs should inform and ask permission from retailers and agree to the most suitable web scraping technique with the retailer's management. Additional measures, such as pauses between "scrapes," may be required to maintain access given that the technology employed to block access may be automated. There may also be legal constraints to web scraping.

5.200 When undertaking web scraping, the same considerations apply as with online collection, most particularly whether the aim is to download from publicly accessible websites, the prices of goods and services sold in

---

[7]Charges relating to a delivery service arranged by the customer for specific items, or for bulk delivery with other items purchased, are legitimate for inclusion in a CPI but should be recorded under a separate heading as indicated in COICOP (2018 COICOP 07.4 Transport services of goods).

[8]Web Robots (also known as Web Wanderers, Crawlers, or Spiders) are programs that traverse the internet automatically. Search engines use them to index the web content. Website owners use the robots.txt file to give instructions about their website to web robots; this is called the Robots Exclusion Protocol. For further information visit http://www.robotstxt.org/robotstxt.html.

physical outlets to increase the efficiency of price collection or whether to download prices paid for online purchases.

5.201 For web scraping to replace traditional price collection, it needs to be demonstrated that prices, both online and in traditional outlets, are the same. While true for some retailers and some products, it is almost certainly not true for all retailers and all products. When prices on the web differ from the prices in outlets, the price collection from a website should be seen as a different outlet type, which should be sampled along with traditional outlets. Traditional price collection would continue along with web scraping. It should be noted that to integrate these prices into a CPI it is necessary to evaluate if there are some extra fees associated with the purchase and that are not included in the prices listed on the website (see paragraphs 5.207–5.209).

5.202 Integration of prices from different data sources needs to consider differences in sampling regimes (for example, the relatively bigger samples facilitated by web scraping) and the relative values of sales. It is one of the reasons why online purchases are often treated as separate elementary aggregates with separate weights.

5.203 Annex 5.6 provides more details on web scraping.

Calculation of Average Price from Different Data Sources in the Elementary Aggregate

5.204 Another aspect to be taken into consideration when combining traditional data sources with web-scraped data refers to the different collection frequencies. Traditional price collection deals with price "snapshots" scheduled in such a way that the price series for a product in an outlet respects both frequency and equidistance with regard to time. In contrast, one of the perceived advantages of web-scraped data is that prices can be collected daily during a certain period, extract the average value and take that as the "snapshot" price for that week. A problem arises when trying to integrate data collected in "continuous daily" frequency into the regular monthly CPI. When using web-scraped data, care must be taken not to apply the raw collected data directly on the calculation of the average monthly price at the elementary aggregate level. Since the number and nature of observations (probably many more than one per month per outlet/period for price obtained by web scraping) are different, they should be transformed into compatible data. This can be accomplished simply by calculating an "outlet monthly price" for the product offer. This average price will then have the same importance as other prices generated by the traditional snapshot approach. The consequences of using directly the online prices will lead to a distorted estimate of inflation, since average prices will be calculated using a disproportionate number of prices that came from online collection, regardless of any relative weight information. Annex 5.5 gives a brief example

of these calculations and how a disregard of the differences in the frequency of price collection can lead to an underestimate of inflation. However, the use of an empirical country-based data set is needed for a more accurate evaluation.

5.205 There is a variety of tools to aid the scraping activity. For example, many programs are available in common programming languages (for example, C, Python, or JavaScript), for standalone applications or, most commonly, as add-ins to the browser. In most cases, web scraping tools are primarily designed to fulfill web application testing and verification. For that reason, most of the tools used are implemented as browser add-ins (or plug-ins). This is not the ideal situation with regard to IT architecture.

5.206 When combining prices using traditional price collection techniques with prices obtained from web scraping, allowance needs to be made in the computation of average product prices for the different collection techniques being deployed, especially the frequency of price collection where web scraping is sometimes carried out at a greater frequency than traditional price collection (see Annex 5.5).

Key Recommendations

- Price collection methods and organization decisions will depend upon country-specific circumstances, available resources, and could potentially vary by item.
- Collected prices should reflect actual transaction prices including any tax and reflecting any discounts, sales, or promotions.
- Items should be priced as often as necessary to ensure that the index reflects a reliable and meaningful measure of price change.
- NSOs should strive to calculate an index based on prices covering the whole period (for example, month or quarter).
- The interval between price observations should be uniform for each outlet.
- The price collection period should be made publicly available and any changes announced well in advance.
- Proper training of price collectors is essential. Detailed documentation on data collection procedures and processes should be drafted and made available to data collectors and CPI staff.
- Quality assurance procedures should be defined and implemented to ensure the accuracy of collected prices.
- Data validation techniques are necessary to ensure the accuracy and reliability of the collected prices.
- Outlier detection methods should be defined and implemented. All questionable prices should be verified and errors are corrected as necessary.

103

Annex 5.1

Consumer Price Index Price Collection Procedures

Figure A5.1   Planning and Organizing Price Collection
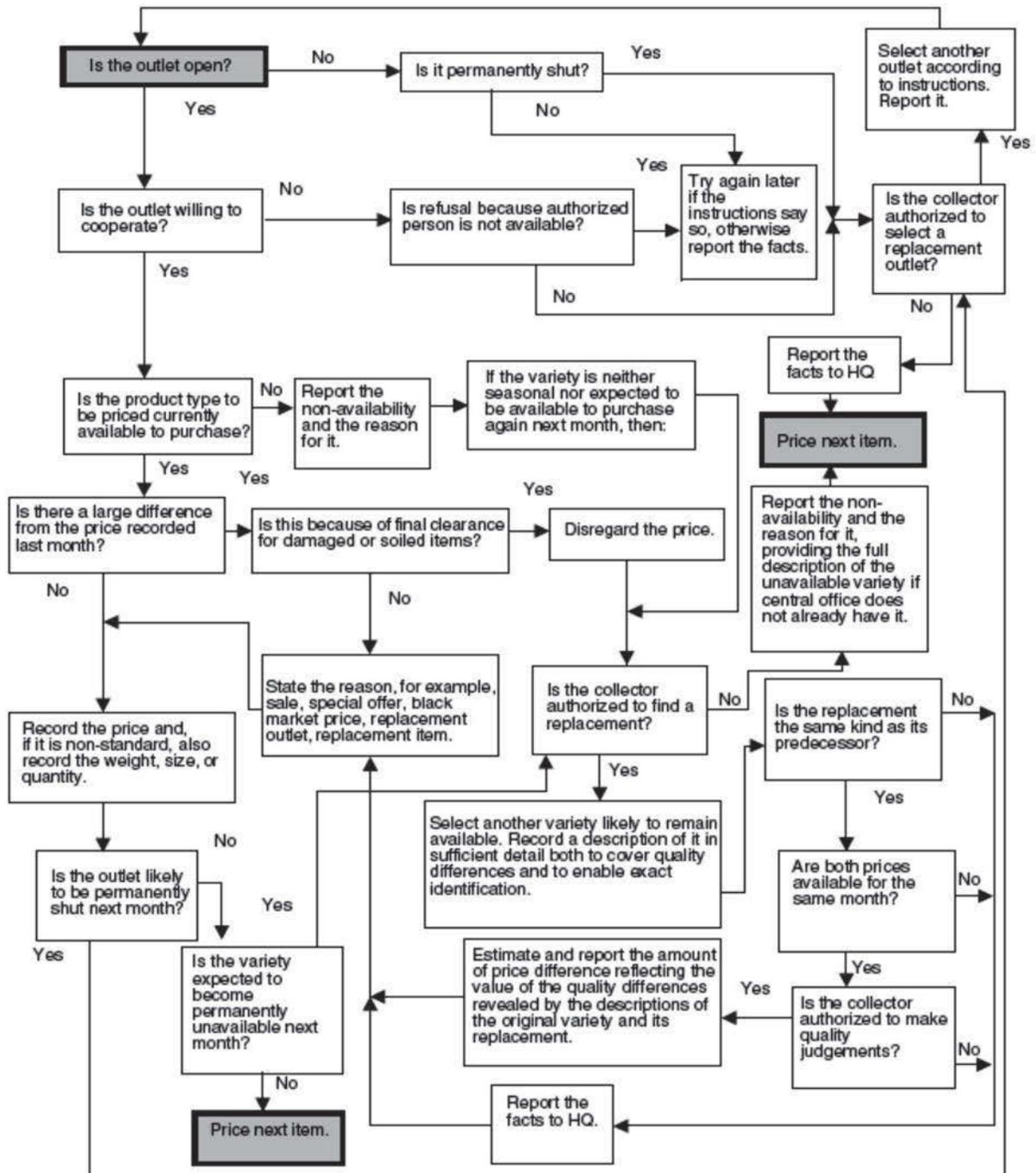
Annex 5.2

Consumer Price Index—Example of a Price Collection Form

Figure A5.2   Price Collection Form

Annex 5.3

Consumer Price Index— Automated Data Checking

Example 1

Example 1 demonstrates the use of median and quartile values to identify outliers. Table A5.1, column A, shows the price ratios for the illustrative sample.

$$S_i = 1 - \frac{R_M}{R_i}, \quad \text{if } 0 < R_i < R_M \qquad \text{(A5.1)}$$

$$S_i = \frac{R_i}{R_M} - 1, \quad \text{if } R_i \geq R_M$$

The first and third quartiles (R_{Q1} and R_{Q3}) and the median (R_M) can be obtained using the quartile function in Microsoft Excel. The average distance of the quartiles from the median (D_M) is defined as

$$D_M = (R_{Q3} - R_{Q1})/2 \qquad \text{(A5.2)}$$

The upper and lower limits are then calculated as

$$L_U = R_M + C \times D_M; \text{ and} \qquad \text{(A5.3)}$$
$$L_L = R_M - C \times D_M$$

Table A5.1 Price Relatives Showing Movement from Previous Period (Example 1)

Price Ratio	S_i	Flagged Using PR	Flagged Using S_i
		Above 1.08723 or Below 0.87628	Above 0.10894 or Below −0.10894
(a)	(b)	(c)	(d)
0.81380	−0.20638	Extreme	Extreme
0.85250	−0.15161	Extreme	Extreme
0.87600	−0.12072	Extreme	Extreme
0.89900	−0.09205	OK	OK
0.90860	−0.08051	OK	OK
0.91350	−0.07471	OK	OK
0.93390	−0.05124	OK	OK
0.94140	−0.04286	OK	OK
0.95530	−0.02769	OK	OK
0.96080	−0.02180	OK	OK
0.96580	−0.01651	OK	OK
0.96680	−0.01546	OK	OK
0.97020	−0.01190	OK	OK
0.97240	−0.00962	OK	OK
0.98170	−0.00005	OK	OK
0.98180	0.00005	OK	OK
0.98430	0.00260	OK	OK
0.98690	0.00525	OK	OK
1.00340	0.02205	OK	OK
1.00500	0.02368	OK	OK
1.00610	0.02480	OK	OK
1.03010	0.04925	OK	OK
1.05710	0.07675	OK	OK
1.08240	0.10252	OK	OK
1.09090	0.11118	Extreme	Extreme
1.09310	0.11342	Extreme	Extreme
1.13000	0.15101	Extreme	Extreme
1.15500	0.17647	Extreme	Extreme
1.22960	0.25246	Extreme	Extreme
1.23040	0.25327	Extreme	Extreme

Table A5.2 Parameters and Derived Limits (Example 1)

Parameter	Series PR	Series S_i
(a)	(b)	(c)
R_{Q1}	0.94488	−0.03907
R_M	0.98175	0
R_{Q3}	1.05035	0.06988
D_M	0.05274	0.05447
C	2	2
L_L	0.87628	−0.10894
L_U	1.08723	0.10894

where the multiplier C is a user-defined value and has been set equal to two to limit the number of observations flagged up as potential errors.

The resulting upper and lower limits are shown in the "Series PR" in column B in Table A5.2.

The price ratio series can be transformed to provide more equal weighting between negative and positive price movements. The transformations are repeated here as

$$S_i = 1 - \frac{R_M}{R_i}, \quad \text{if } 0 < R_i < R_M \qquad \text{(A5.4)}$$

$$S_i = \frac{R_i}{R_M} - 1, \quad \text{if } R_i \geq R_M$$

The transformed observations are shown in the "Series S_i" in column C in Table A5.1. The quartiles, median, and calculated limits for the transformed series are shown in column C in Table A5.2. The increased value for D_M for the transformed sample shows that the transformation has increased the distances for the price decreases while leaving the distances for positive movements the same.

Columns C and D in Table A5.1 show, respectively, the observations that would be flagged for further examination (indicated by the word "extreme") for the original price ratios and the transformed price movements.

Example 2

Example 2 demonstrates the same statistical filtering method as in Example 1, but with 16 additional price ratios added to the sample. All the new price ratios show zero change. The same calculations are done but on a sample of 46 instead of 30 observations. Table A5.3 shows the sample of price ratios and the transformed price movements, as well as the observations flagged for further observations. Table A5.4 shows the parameters and calculated limits.

A comparison of the results from the two examples demonstrates the effect of having a significant number of observations with no price movement. The distance from the median (D_M) is reduced and the number of observations flagged for further examination is significantly increased.

Example 3

Example 3 demonstrates the alternative statistical filtering method, the Tukey algorithm. The enlarged sample from Example 2 is used here to demonstrate the benefit of this method when the sample has a large proportion of price ratios indicating no movement. Table A5.5 presents the intermediate

109

Table A5.3 Price Relatives Showing Movement from Previous Period (Example 2)

Price Ratio	S_i	Flagged Using PR	Flagged Using S_i
		Above 1.03490 or Below 0.96510	Above 0.03598 or Below −0.03598
(a)	(b)	(c)	(d)
0.81380	−0.22880	Extreme	Extreme
0.85250	−0.17300	Extreme	Extreme
0.87600	−0.14160	Extreme	Extreme
0.89900	−0.11230	Extreme	Extreme
0.90860	−0.10060	Extreme	Extreme
0.91350	−0.09470	Extreme	Extreme
0.93390	−0.07080	Extreme	Extreme
0.94140	−0.06220	Extreme	Extreme
0.95530	−0.04680	Extreme	Extreme
0.96080	−0.04080	Extreme	Extreme
0.96580	−0.03540	OK	OK
0.96680	−0.03430	OK	OK
0.97020	−0.03070	OK	OK
0.97240	−0.02840	OK	OK
0.98170	−0.01860	OK	OK
0.98180	−0.01850	OK	OK
0.98430	−0.01600	OK	OK
0.98690	−0.01330	OK	OK
1.00000	0.00000	OK	OK
1.00000	0.00000	OK	OK
1.00000	0.00000	OK	OK
1.00000	0.00000	OK	OK
1.00000	0.00000	OK	OK
1.00000	0.00000	OK	OK
1.00000	0.00000	OK	OK
1.00000	0.00000	OK	OK
1.00000	0.00000	OK	OK
1.00000	0.00000	OK	OK
1.00000	0.00000	OK	OK
1.00000	0.00000	OK	OK
1.00000	0.00000	OK	OK
1.00000	0.00000	OK	OK
1.00000	0.00000	OK	OK
1.00000	0.00000	OK	OK
1.00340	0.00340	OK	OK
1.00500	0.00500	OK	OK
1.00610	0.00610	OK	OK
1.03010	0.03010	OK	OK
1.05710	0.05710	Extreme	Extreme
1.08240	0.08240	Extreme	Extreme
1.09090	0.09090	Extreme	Extreme
1.09310	0.09310	Extreme	Extreme
1.13000	0.13000	Extreme	Extreme
1.15500	0.15500	Extreme	Extreme
1.22960	0.22960	Extreme	Extreme
1.23040	0.23040	Extreme	Extreme

Table A5.5 Price Relatives Showing Movement from Previous Period (Example 3)

Price Relative	Less 5 Percent Tails and Less Zero Movement	T Lower Set	T Upper Set	Flagged as Extreme
				Above 1.17184 or Below 0.88332
(a)	(b)	(c)	(d)	(e)
0.8138				Extreme
0.8525				Extreme
0.8760	0.8760	0.8760		Extreme
0.8990	0.8990	0.8990		OK
0.9086	0.9086	0.9086		OK
0.9135	0.9135	0.9135		OK
0.9339	0.9339	0.9339		OK
0.9414	0.9414	0.9414		OK
0.9553	0.9553	0.9553		OK
0.9608	0.9608	0.9608		OK
0.9658	0.9658	0.9658		OK
0.9668	0.9668	0.9668		OK
0.9702	0.9702	0.9702		OK
0.9724	0.9724	0.9724		OK
0.9817	0.9817	0.9817		OK
0.9818	0.9818	0.9818		OK
0.9843	0.9843	0.9843		OK
0.9869	0.9869	0.9869		OK
1.0000				OK
1.0000				OK
1.0000				OK
1.0000				OK
1.0000				OK
1.0000				OK
1.0000				OK
1.0000				OK
1.0000				OK
1.0000				OK
1.0000				OK
1.0000				OK
1.0000				OK
1.0000				OK
1.0000				OK
1.0000				OK
1.0034	1.0034		1.0034	OK
1.0050	1.0050		1.0050	OK
1.0061	1.0061		1.0061	OK
1.0301	1.0301		1.0301	OK
1.0571	1.0571		1.0571	OK
1.0824	1.0824		1.0824	OK
1.0909	1.0909		1.0909	OK
1.0931	1.0931		1.0931	OK
1.1300	1.1300		1.1300	OK
1.1550	1.1550		1.1550	OK
1.2296				Extreme
1.2304				Extreme

Table A5.4 Parameters and Derived Limits (Example 2)

Parameter	Series PR	Series Si
(a)	(b)	(c)
R_{Q1}	0.96765	−0.03343
R_M	1.00000	0
R_{Q3}	1.00255	0.00255
D_M	0.01745	0.01799
C	2	2
L_L	0.96510	−0.03598
L_U	1.03490	0.03598

data stages in addition to the basic sample and the indicator flagging extreme observations as possible errors.

The price ratios for the sample are shown in column A. The first step was to remove the highest and lowest 5 percent of price ratios. Five percent of this sample equals 1.5 observations. This was rounded up to two observations so the two highest and the two lowest price ratios were removed. Observations with zero price movement were also removed. The remaining observations are shown in column B. The arithmetic mean (AM) of the remaining set of observations was calculated. This value, along with other parameter calculations is shown in Table A5.6. The arithmetic means of the lower and upper sets of data are then calculated (labeled AM_L and AM_U, respectively). The lower and upper data sets have been presented in columns C and D, respectively, of Table A5.5 purely for explanatory

110

Table A5.6   Parameters and Derived Limits (Example 3)

Parameter	Value
(a)	(b)
AM	0.99429
AM_L	0.94990
AM_U	1.06531
T_L	0.88332
T_U	1.17184

purposes. The Tukey lower and upper limits are then calculated as

$$T_L = AM - 2.5(AM - AM_L) \qquad (A5.5)$$
$$T_U = AM + 2.5(AM_U - AM)$$

The results are shown in Table A5.6.

Using this method, five observations would be selected for further examination—many fewer than the 18 selected in Example 2.

111

Annex 5.4

Documentation Control Template

The documentation control template is an essential element of documentation, production, and dissemination and control. Documentation control contributes to better quality management as access to nonauthors is restricted to "read only." It provides checks, background (including explanations for changes), and an audit trail. Two further benefits accrue when combined with an electronic system:

- More efficient production of documentation as it helps with initial compilation and updates, and reduces the need to print and circulate paper copies
- Better informed staff because they have immediate electronic access to the latest documentation, including desk instructions, with search facility by subject and author

Figure A5.3 Documentation Control Template

Date Issued	Documentation	Reference	Details of Change	Reason for Change	Name of Issuer
Day/Month/Year	Calculating CPI Food Item Weights (nonseasonal)	2.1	Change in Process with Effect from . . . (date)	CPI Technical Board Has Agreed That in Future Weights Should Be Taken from National Accounts	L. Smith, Statistician, CPI Program
XX/XX/XX	Calculating and Updating Price Index for Telecommunication Services	2.5	Change in Process with Effect from . . . (date)	Methodology Changes in Pricing Structures for Mobile Telephones—New Methodology Agreed by CPI Technical Board. Reflects Changing Market	L. Smith, Statistician, CPI Program
XX/XX/XX	Desk Instructions for Checking and Editing of Prices	3.1	Additional Checks to Be Carried Out Based on Month-on-Month Price Change	Last Audit Indicated Current Checks Inadequate Resulting in Incorrect Prices Entering the CPI	C. Brown, Operations Manager, CPI Program

Annex 5.5

The Calculation of Average Product Price When Combining Prices from Different Price Collection Methods and for Different Price Collection Frequencies

When combining traditional and web scraping price collection methods, there are two options for average product price computation. Web scraping can follow the same calendar collection that is used in traditional collection. With this option, the advantages of web scraping are disregarded. The second option will be to collect prices more frequently with web scraping techniques. This will increase the number of observations used in the computation, potentially providing more reliability to the estimation. In the following simplified example relating to just one item, the price for a specific product A is collected in a sampled local outlet in the framework of traditional price collection twice along the considered time frame of 10 days. The price for the same product is collected on a daily basis using web scraping. These data are collected for two months and the price at day 1 and at day 10 is the same for both outlet/collection types. However, the price averages and the indices

will differ according to the way these are computed at the elementary aggregate level.

There are two methods of calculating the average price (geometric mean) at the elementary aggregate level. The assumption is made that there are no expenditure weights (and each observation at a given point in time has the same weight) and prices are collected to be representative of all sales.

- Method 1: compute the geometric mean in two steps: first, an average price is calculated by kind of outlet/collection method; second, a geometric average is calculated with the two average prices per outlet
- Method 2: compute the geometric mean in a single stage using all prices (not recommended)

Taking into consideration the different collection frequencies, method 2 will undervalue price change (noting that in this example prices are falling) since it takes into account a huge number of observations taken every day where the price change is not significant and it is providing more "weight" to the internet outlet with a price series that is more stable than to the local outlet where the price being observed each 10 days is recording a higher change. Due to these differences and the circumstance described, method 1 is recommended since it will provide the same "weight" to both kinds of outlets/collection methods. In summary, the prices can only be averaged in a single step when there is an equal number of price observations generated by each price collection for each item over a given period. Averaging prices in a single step when the frequency of price collection varies between different price collection methods leads to distorted results.

Table A5.7 Combining Prices from Different Price Collection Methods and for Different Price Collection Frequencies

Product A	Month 1										Geometric Mean	
	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10		
Local Shop (traditional collection)	500									480	489.898	
Internet Website (web scraping)	500	500	500	497	497	490	489	483	480	480	491.536	
Geometric Mean: Method 1											490.717	
Geometric Mean: Method 2											491.263	
Product A	Month 2										Geometric Mean	Indices
	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10		
Local Shop (traditional collection)	450									430	439.886	89.79
Internet Website (web scraping)	450	452	452	448	445	445	445	440	435	430	444.146	90.25
Geometric Mean: Method 1											442.011	90.02
Geometric Mean: Method 2											443.433	90.17

Annex 5.6
Web Scraping

Introduction

The collection of prices is an integral component of the production of a CPI. Several data collection modes are available and used by countries. These include personal visits, online, telephone, administrative data, and transactions data. More recently, with the growth of online retailing, pricing information may be obtained directly from websites. Advances in technology and automated scraping software have enabled large-scale data collection from the internet. This is referred to as web scraping.

Benefits of Web Scraping

Web scraping enables many more products to be priced, and for these products to be priced more often than would be possible using traditional data collection methods. Web scraping provides an opportunity to significantly enhance the sample of products and prices collected by expanding product coverage. Other benefits of web scraping include:

- Automated data collection at a reduced collection cost.
- Enhanced price representativity.
- Increased price collection frequency (that is, daily versus once per month) allows for a more representative price for the period to be obtained. Secondary effects of using an average period price include reduced volatility.
- Faster identification of new and disappearing products.
- Reduced respondent burden.
- Rich source of metadata (that is, product characteristics) that can be extracted, stored, and potentially used for an explicit quality adjustment (for example, hedonics). Such metadata may also complement other data sources such as transactions data.

Limitations of Web Scraping

The lack of expenditure information from web scraping means that products/product groups cannot be weighted by economic importance and limits the types of index formulas that can be used to compile the index. Other potential limitations of web scraping are the following:

- Web scraping is limited to retail outlets that have an online presence (that is, potential for undercoverage).
- Web scraping requires regular IT maintenance. Website changes may cause the web scraping program to fail. Businesses can block Internet Protocol addresses if they detect the web scraping activity and wish to prevent it.
- If performed within the NSO, web scraping requires compilers with intermediate programming knowledge to deal with the regular IT maintenance.
- Web scraping distinct prices for different geographic regions may be difficult if the website detects the physical location of the computer's Internet Protocol address. This may require assumptions that retailers use national pricing for regional price indices.

- Web scraping may be time-consuming for large retailers with many different webpages and products.

The Process of Web Scraping

For those NSOs using web-scraped data for research and production purposes, the process of performing web scraping has focused on two main methods:

- Web scraping performed within the NSO using statistical software[9]
- Web scraping services procured from a third-party/private company[10]

The choice to perform web scraping within the NSO or contracting with an external vendor will depend on the local context in which the NSO operates (for example, budgets, NSO programming capacity, maintenance costs). With respect to those NSOs performing web scraping internally using statistical software, the process of web scraping information from the internet can be broken down into three main steps:

- Confirm if the website allows scraping
- Scrape the website
- Clean the collected data

There are two main ways to determine whether a website is eligible for web scraping. First, staff setting up web scrapers should check the terms and conditions section of a website for "conditions of use." Here, websites will often specify if web scraping is allowed or prohibited. Additionally, a "robots.txt" file can be located within the root directory of a website. These text files contain detailed information and may outline the conditions for web scraping possibly including who is allowed to web scrape, which information is available for scraping, and anything for which scraping is forbidden.

Once it is determined that a website is available for scraping, a scraper is set up for the website. Each website is unique and the optimal scraping strategy may change from one site to the next. The scraper locates the website's category structure and identifies all the relevant categories to be scraped. The programmer defines the parts of the structure to be included and excluded. For example, a website may list all the individual product categories, then additional categories such as "new products" or "all products" which duplicate the products in the individual categories. These additional categories can be excluded by the programmer.

The scraper then proceeds to download all products and prices from the internet. An attempt is made to show as many products on each page as possible by experimenting with URL options on the website prior to setting up the scraper. There are two options available for pulling in the information. First, data may be collected as text that essentially uses scrapers to copy and paste from the website. In this case, the process of cleaning the text is carried out following the scraping. Alternatively, products and prices can be pulled in using designated HTML tags and classes which provide a more targeted approach to extracting and cleaning the data. Another advantage of this approach is that product identifiers

---

[9]Van Loon and Roels (2018).
[10]Bentley and Krsinich (2017).

Table A5.8 Web Scraping—Typical Data Structure

Date	Retailer	Category	Product ID	Price
July 10, 2019	Retailer ABC	Children's Shirts	Brand XYZ—Short Sleeve Polo Shirt	$45.00
July 10, 2019	Retailer ABC	Children's Shirts	Brand XYZ—S/S Regular Shirt	$55.00
July 10, 2019	Retailer ABC	Children's Shirts	Brand XY—Short Sleeve Regular T-Shirt	$15.00
July 10, 2019	Retailer ABC	Children's Shirts	Brand XYZ—Long Sleeve Regular Shirt	$65.00
July 10, 2019	Retailer ABC	Children's Shirts	Brand XYZ—Short Sleeve Regular Shirt	$35.00

can occasionally be hidden in the HTML so pulling them in using the tags allows these to be added to the product description (as opposed to relying purely on the text description). However, the use of HTML tags is not easy for every website.

For information collected as text, the raw data need to be cleaned post-scraping so that only the set of products and prices remain. A pattern in the data needs to be uncovered and coded to separate the products and prices from the "noise" (including removing all the information prior to and after the list of products). For products which are on sale, multiple prices may be listed. In these instances, the scraper records the sale prices as the current price of the product.

With regard to the available information to construct price indices, Table A5.8 provides a summary of a typical metadata scraped by NSOs. In summary, the data frame will typically include:

- Date: specific day of the scrape (date)
- Retailer: name of the retailer (text)
- Category: retailer's website classifications (text)
- Product ID: text description of product (text)
- Price: specific price of product (numeric)

Practical Considerations

The basic information required to compile a price index includes prices, expenditure information (or reasonable assumptions on substitution if no expenditure information is available), and classifications (for both products and product groups). While web-scraped data (such as the example in Table A5.8) may appear reasonably consistent with these requirements, it is important to ensure the conceptual requirements of the CPI (for example, quality adjustment) are maintained. "Big data (transaction, online, and administrative data) is 'found data' in the sense that measuring CPI inflation is a secondary use of the data—the data were not created with this use in mind."[11]

This subsection describes some of the main challenges identified by NSOs when using web-scraped data, including:

- Classifying web-scraped data
- Options to define individual products
- Index aggregation options

[11] Bentley and Krsinich (2017, 6).

Classification of Web-Scraped Data

The classification of web-scraped data involves similar considerations as described in Chapter 10 for the use of scanner data. Web-scraped data typically have some basic product text and category description that are required to be mapped to a NSO hierarchical classification (for example, COICOP). Approaches considered by NSOs to solve these classification problems include:

- *Text string searches*: Check for the presence or absence of keywords in the description string for classification.
- *Category mapping*: Some of the data sets (or parts thereof) contain retailer categories for each product; if one of these categories sits within a classification, the category can be mapped to the classification.
- *Manual mapping*: A compiler looks at the description string. This is the most feasible option for small data sets.
- *Supervised learning algorithms*: Provide training data (for example, using one or multiple methods previously mentioned) to statistical learning algorithm that identifies patterns between text and training decision for automatic classification.

Options to Define Individual Products

An essential part of price measurement is accounting for quality change and the introduction of new products. The CPI measures the price inflation in a basket of goods and services priced at constant quality. If the quality of a product changes over time, then prices are adjusted so that the index movements reflect pure price change. This has important implications for web scraping.

The appearance and disappearance of products from a CPI sample has the potential to bias the index unless any corresponding changes in the quality of the sample are dealt with appropriately. This poses a problem for the calculation of indices incorporating all (or most) web-scraped prices due to a large number of prices these data sets contain, the high rate of product attrition, and the tendency for products to have unusual price movements near the start and end of their life cycles.

One approach for dealing with this problem is to estimate the price change between two periods using the products available at both time points only, thus excluding the prices of new and disappearing products. The matched-model method (as described in Chapter 6) involves discarding information about new and disappearing products. However, it gains strength from the comprehensive coverage from the census of products represented in the full web-scraped data set. This can be considered as an application of the overlap method, as it is almost certain that sales of new and disappearing products would overlap with the sales of other products that consumers would use as substitutes.

An assumption behind the overlap method is that price differences between products are reflective of quality differences. This assumption seems reasonable in a competitive marketplace and in normal circumstances. However, disappearing products are sometimes sold at discounted prices to clear remaining stock (end of life cycle), and if not linked to a product of comparable quality, may produce a "relaunch" problem and could potentially create a downward bias in the

index.[12] This problem has been identified in various price index studies on different types of products including high-technology goods,[13] clothing,[14] and personal care products.[15]

To overcome this problem, several practical strategies have been proposed, primarily focused on extracting characteristics information (for example, brand and shirt type) from text strings to form broader product definitions. Key techniques proposed are the following:

- Use of a broader product category (for example, children's shirts in Table A5.8)
- Text/regular expression functions: used to extract characteristics from semistructured text data (for example, "XYZ" in Table A5.8 extracts the characteristic "brand")
- Approximate (fuzzy) matching functions: used to approximately match text strings using a penalty function
- Supervised learning algorithms:[16] provide training data to statistical learning algorithm that identifies patterns between text and training decision for product classification
- Unsupervised learning algorithms:[17] use characteristics (for example, text string and price) and algorithms to automatically define "clusters" of products

Options to Aggregate Prices

Web-scraped data do not contain quantity or expenditure information. This would naturally restrict the choice of index formula to an unweighted price index formula (for example, Jevons), which is the current choice for many NSOs in the CPI production.

The use of web-scraped data in compiling price indices continues to be the subject of extensive research. Some researchers have experimented with approximating expenditure weights based on data observed from the websites.[18] Using brand and product-type definitions and the number of products as a proxy for quantity data, results demonstrate that the web-scraped data approximate a benchmark index (scanner data) using the Geary–Khamis method. Additional published studies (for example, Metcalfe and others [2016]) experiment with several bilateral and multilateral price index methods. Research findings showed a substantial

amount of drift across most bilateral and multilateral indices with granular product definitions for clothing products.

Country Case Study—Web Scraping

As part of a broader project to modernize the CPI, one NSO began collecting prices using web scrapers beginning in May of 2016. Web scrapers are currently programmed and maintained by the NSO staff using Microsoft Excel (Visual Basic for Applications), collecting approximately 500,000 prices per week across more than 50 retailers. Web scraping increases the sample of prices used to compile the index, thus providing a more meaningful measure of price change. From the second quarter of 2017, the NSO began using web-scraped data in the calculation of the CPI. Varieties are selected using the same methods as other forms of price collection (field collection, online collection, or other methods). An average price is calculated for each item over a given period.

Representative and stable products are selected. When transitioning a respondent from using traditional data collectors to web-scraped prices, an attempt is made to link the current field-collected product to the identical product on the website. A determination regarding quality adjustment is made for each product to ensure that the new web-scraped product's base period price is correct.

For each web-scraped product in the sample, the price for a given period (month or quarter) is an arithmetic average of the prices which fall within the specified period. If a product disappears, a replacement product is selected from among the other products within the relevant category from that respondent. A quality adjustment is performed to ensure that only pure price change is shown. The combination of category name, brand/product description, and price history of both the old and new product is sufficient to enable an accurate quality adjustment to be applied.

This NSO has adopted a phased approach to implementing new retailers and price index methods using web-scraped data. With respect to respondents, an assessment is made for each respondent to be transitioned to web-scraped data, taking into account the quality of web-scraped data over a period of time, the correlation between online prices and field-collected prices in each city, and the potential for collection efficiencies and sample improvements. With respect to price index methods, development work continues within the NSO on text mining techniques (to form broader product definitions, especially for clothing) and price index methods (both bilateral and multilateral index methods) that maximize and automate the use of web-scraped data.

---

[12] Chessa (2016).
[13] Silver and Heravi (2005).
[14] Chessa (2017).
[15] Chessa (2016).
[16] Abe and Shinozaki (2018).
[17] Metcalfe and others (2016).
[18] Chessa and Griffioen (2016).