Sample size for the estimate of consumer price sub-indices with alternative statistical designs [1]

Carlo De Gregorio [*]

Abstract

This paper analyses the sample sizes needed to estimate Laspeyres consumer price sub-indices under a combination of alternative sample designs, aggregation methods and temporal targets. In a simplified consumer market, the definition of the statistical target has been founded on the methodological framework adopted for the Harmonized Index of Consumer Prices. For a given precision level, sample size needs have been simulated under simple and stratified random designs with three distinct approaches to elementary aggregation, founded on Carli, Jevons and Dutot formulae. Alternative temporal targets are also examined: the single monthly target, the whole sets of monthly and quarterly indices, the annual average and the annual link. Empirical evidence is finally provided, based on the elaboration of survey microdata referred to elementary aggregates - such as air transport and package holidays - characterized by high volatility within and between months.

Keywords: chained index; aggregation; sample size; sampling variance; air transports; package holidays.

1. Introduction

In recent years there has been a growing concern for a more explicit use of the concepts and tools of statistical inference to produce estimates of consumer price indices (CPI) and, in particular, to define the targets of the estimates in the fashion typical of statistical survey methods. Although this issue has never been at the core of CPI literature, the pioneering works on this subject date back to Banerjee (1956) and Adelman (1958), while systematic research on the sampling variance of the Laspeyres CPI index has been developed since mid-eighties: see for example the session dedicated to this issue at the 1987 joint ISI-IASS

[1] I wish to thank Alexandre Makaronidis, former head of the HICP unit at EUROSTAT, Keith Hayes (head of HICP methodology) and Jan Walschots, as well as the colleagues of the National Statistical Institutes participating to the EUROSTAT Task force on HICP sampling. I am also grateful to my colleagues at ISTAT: Paola Anitori for her helpful comments, criticism and suggestions, and Alessandro Brunetti for further observations. I am as well grateful to the other colleagues of the former CPI unit, namely Carmina Munzi (this paper is dedicated to her memory), Patrizia Caredda, Stefania Fatello, Rosanna Lo Conte, Maurizio Massaroni, Stefano Mosca, Francesca Rossetti and Paola Zavagnini. I finally acknowledge the useful comments and advices of an anonymous referee. Nevertheless, the views expressed here are solely those of the author and do not necessarily reflect those of ISTAT nor EUROSTAT. An earlier draft of this paper has been presented at the 12th meeting of the Ottawa group (International working group on price indices), held in Wellington (NZ) the 4-6 May 2011.

[*] Senior researcher (Istat), email: cadegreg@istat.it

conference and in particular the works of Biggeri et al. (1987), Andersson et al. (1987) and Leaver et al. (1987), or the works by Kott (1984) and more recently by Dalèn (2001).[2] Further research focused also on specific issues, such as the properties of sample designs based on alternative formulas for elementary aggregation (Dalèn 1992; Baskin et al. 1996; Fenwick 1998; ILO 2004, chap.5; Balk 2008, chap. 5). Stimuli for the adoption of more developed statistical techniques also came along with the innovations in price collection, especially in selected consumer markets: this happened with scanner data (De Haan et al. 1997; Fenwick 2001; Koskimäki et al. 2003) and with sources like e-commerce and administrative or private databases. In general, the availability of larger and more flexible data sets of price quotes made it necessary to set up generalized methods, fostering a greater attention on sampling issues. The integration with other statistical sources also favoured the adoption of sampling based approaches: for instance, the availability of regularly updated business registers has been considered to improve sample design in Biggeri et al. (2006).

Nevertheless, most of the empirical approaches adopted to measure the variance of the estimates relied on the use of replication techniques, since the data available for analysis derive mainly from purposive samples and quite rarely from probabilistic designs. Several authors dealt with this issue (Biggeri et al. 1987; Andersson et al. 1987; Balk 2008, p.176) and with the need to provide a suitable statistical design (Kott 1984; Dalèn et al. 1995; Dalèn 1998, 2001; Ribe 2000; Dorfman et al. 2006). The definition of both the universe and the target parameters appears by far the most critical issue in a CPI sampling design and, more in general, in the CPI itself (Dalèn 1998, 2001). Several factors connected to the rapid evolution of consumer markets impair this definition: they are related to products and outlets replacements as well as to the changes in their characteristics, and they should be tackled, at least theoretically, in order to provide a solid foundation for the production of the estimates. Therefore, the need of a structured framework of concepts and definitions has emerged in order to reduce complexity, and to provide sufficiently general and operative solutions.[3]

Ribe (2000) and the most recent methodological developments of the chained Laspeyres index adopted in the EU Harmonized Index of Consumer Prices (hereafter HICP) propose to structure the universe of transactions into homogeneous partitions based on the concepts of product-offer and consumption segment. Starting from this approach, this paper proposes a definition of the target universe under some assumptions on the functioning of consumer markets (they help to simplify the identification of the statistical target), and provides a tool to measure the sample size needed to estimate the sub-indices of HICP under alternative designs. The results obtained with simple and stratified random designs are compared, taking into account the use of alternative criteria for segmentation, elementary aggregation and temporal targets. Two case studies are also developed.

Based on the annually chained Laspeyres formula used in the HICP, a definition of the statistical target for a monthly index is firstly provided (par. 2). Given the desired precision

2 The literature on this subject is briefly surveyed in Dalèn et al. (1995). See also Wilkerson (1967), Dippo et al. (1983), Leaver et al. (1987), Leaver et al. (1991), Baskin et al. (1996), Norberg (2004). For an overview of variance estimation approaches in selected countries see ILO (2004, chap.5).

3 “*The consumer market ultimately consists of an enormous (but finite!) number of transactions, where goods and services (products) are purchased by consumers. However, it is not feasible to compare transactions directly between periods. Like the physicists who divide matter successively into molecules, atoms and nucleons, we have to bring some structure into our market universe as a prerequisite for a measurement procedure.*” Dalèn (2001, p.3).

level, we derive the sample size with simple and stratified random designs under Carli, Jevons and Dutot aggregation. For each design, alternative temporal targets are examined such as monthly, quarterly and annual indices, as well as the annual link of the chaining sequence (par. 3). The approaches are then tested on an experimental ground, simulating artificial populations from the microdata relating to two sub-indices of the HICP - air transports and package holidays - both characterized by a high volatility of price dynamics within and between months (par. 4).

2. The statistical target

2.1. Some aspects of the construction of the HICP

The HICP is a monthly Laspeyres index based on the average of a reference year (*yr*).[4] It is built as a chained index by linking together the monthly price indices *I* of the current year *y* based on the link month of *December y-1* and the fixed base index *H* of *December y-1*. By iteration, in the reporting month *m* of year *y* the aggregate HICP is derived as the product of three elements: a fixed base index (*H*), the product of the *(y-yr-1)* annual links, and the link index of the reporting month (*m*). In formulas:[5]

$$H^{y,m} = H^{y-1,12} I^{y,m} = H^{yr,12}\left(\prod_{x=yr+1}^{y-1} I^{x,12}\right) I^{y,m} \quad (1).$$

The link of the reporting month $I^{y,m}$ is compiled as the weighted average of the sub-indices referred to an exhaustive set of disjoint aggregates *j* of the total consumer expenditure in the weight reference year:

$$I^{y,m} = \sum_j I_j^{y,m} w_j^y \quad (2)$$

where the expenditure weights *w* add up to unity and in principle change every year since they are referred to the consumption expenditure of year *y-1*.[6]

The construction of any HICP aggregate follows a hierarchical procedure: the aggregation of the link indices comes first, then the result is chained to the fixed base index of the same aggregate. The sub-indices $I_j^{y,m}$ can be therefore interpreted as the primary components of the HICP and, as a consequence, each of them represents a distinct statistical target (Ribe 2000, p.1): hereafter we shall refer to the problem of estimating these sub-indices. Notice also that expression (2) can be applied to any exhaustive partition of the target consumption expenditure: we choose in particular to deal with the partition realized through the groups of COICOP-HICP classification at the lowest level of detail used for

---

[4] At present yr=2005.

[5] See EUROSTAT (2001, p.175-197). To simplify notation, in what follows the basis of all indices has been set to 1 instead of the usual 100.

[6] Furthermore, the weights are price-updated from the weight reference period (y-1) to the price reference period (December y-1). See EUROSTAT (2001, p.188-190), Hansen (2006), ILO (2004, chap.9).

HICP dissemination. This partition concerns almost 100 sub-indices (EUROSTAT 2001, p.253-68).

2.2. Re-pricing of transactions and statistical target in a HICP perspective

One of the main advancements in HICP methodology and legal basis regards the statistical definition of the HICP universe.[7] In particular, the target parameter for the annual links of a monthly Laspeyres CPI corresponds to the ratio of two simulated consumption expenditures obtained by mapping the universe of transactions in the weight reference period (year *y-1*) onto the sets of available offers in the price reference period (*December y-1*) and in the reporting period (month *m* of year *y*). In order to define this *re-pricing* of transactions, the concept of product-offer was introduced in EC Regulation 1334/2007: "*product-offer means a specified good or service that is offered for purchase at a stated price, in a specific outlet or by a specific provider, under specific terms of supply, and thus defines a unique entity at any one time"*. As a matter of fact, product-offers are the observation units in CPI sampling and they determine the partition of total transactions. Nevertheless they represent a rapidly changing stock: they may change as the characteristics of the goods and services evolve, as they are replaced, as retail evolves, or simply as prices change. In order to provide stable entities on which price comparisons can be based, the sets of all the transactions and product-offers in the statistical universe are exhaustively clustered into consumption segments, where each segment identifies homogeneous product-offers with regard to marketing targets, consumption purposes and characteristics. In the HICP framework, consumption segments represent the fixed objects to be followed by the Laspeyres index.[8].

The mechanism of re-pricing can be summarised on the basis of Figure 1, in which full information is assumed on consumer expenditure. The squared area on the left summarises the total consumer expenditure in the weight reference year *y-1* within a given consumption segment: the geometric shapes (circles, lozenges and hexagons) identify the product-offers, while the black smaller circles represent the actual transactions. Only a subset of the product-offers and transactions of year *y-1* is in common with the price reference month (that is *December y-1*) – the upper central side of Figure 1. Given this information, the consumer expenditure in year *y-1* is simulated by means of mapping functions connecting the product-offers available in that year with those available in the price reference month.[9] An identical approach is adopted for the re-pricing based on the reporting month (*y, m*): in this case there is no overlapping with transactions and product-offers in year *y-1*. The index

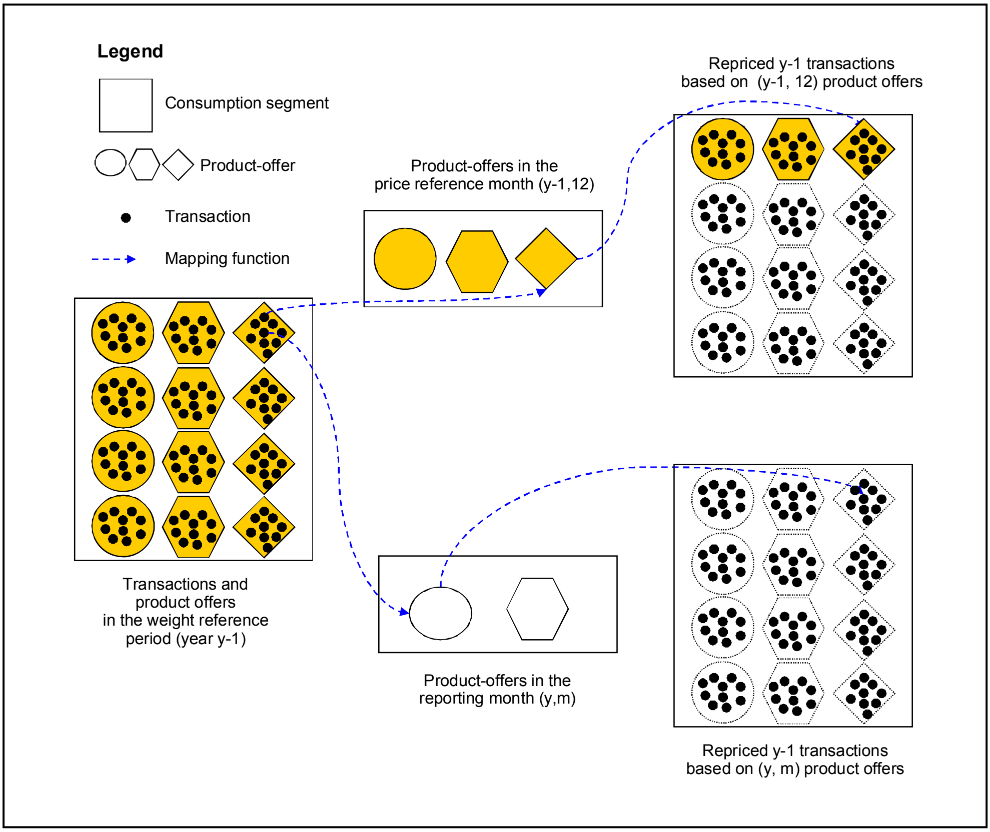
7 See both Ribe, 2000 and Commission Regulation (EC) no. 1334/2007; a collection of the early HICP legislation can be found in EUROSTAT (2001).

8 In particular: "*(...) 'consumption segment' means a set of transactions relating to product-offers which, on the grounds of common properties, are deemed to serve a common purpose, in the sense that they: are marketed for predominant use in similar situations, can largely be described by a common specification, and may be considered by consumers as equivalent. (….). The notion of consumption segments by purpose is therefore central to sampling and to the meaning of quality change and quality adjustment. However, an ambiguity in this concept concerns the level of aggregation at which it is defined and applied. (...) The range of product-offers will change over time as products are modified or replaced by retailers and manufacturers. The HICP requires the representation of all currently available product-offers within the consumption segments by purpose selected in the reference period in order to measure their impact on inflation. This applies particularly to new models or varieties of previously existing products.*" (EC Regulation 1334/2007).

9 The common set of *December y-1* transactions is simply replicated (see the continuous border forms in the first row of the upper right square in Figure 1) while the rest of *y-1* expenditure is simulated (dotted border).

for the consumption segments is finally obtained as the ratio of the value of the two sets of simulated transactions.

Figure 1 - The re-pricing of transactions



Re-pricing has the advantage of providing a general framework for the provision of suitable solutions for the statistical treatment of inflation estimates. Mapping functions implicitly or explicitly incorporate various aspects of consumer behaviour modelling, and consequently define statistical imputation techniques, non response treatment, quality adjustment. They represent the methodological core of the estimates: their complexity directly depends on the rapidity of the changes that occur in the set of product-offers both generated by changes in the price level and by the range of the goods supplied to consumers. It is easily understandable that mapping functions are open to host several alternative hypotheses. As a matter of fact, the whole framework for the definition of HICP statistical universe is a theoretical tool open to a wide range of possible solutions, while methodological and empirical research is still needed in order to test its applicability as a statistical tool. Another key point is given by the definition of consumption segments. HICP regulation itself recognizes that ambiguities still concern the level of aggregation with which consumption segments are defined and applied. It is likely that consumption segments need to be specified case-by-case and this fosters the strategic role of consumer markets analysis, such as for example the structure of supply and demand, the marketing approaches and the segmentations adopted by producers and dealers. This part of the job strictly interacts with the definition of mapping functions.

2.3. The definition of the statistical target

Given the Laspeyres formula, and following the approach set up in Ribe (2000) and in HICP legal basis, the starting point for the definition of each target sub-index is given by the set of all the transactions in the weight reference year *y-1* concerning the COICOP-HICP group *j*. We assume a perfect knowledge of all information necessary to compile the indices. In particular, each transaction in the weight reference period is tracked; it concerns the purchase of a product-offer, and each product-offer is attributed to a specified consumption segment.

Product-offers are defined by the combination of two sets of characteristics. A first set consists of a vector g_i of variables describing the product, the outlet and the corresponding consumption segment (*h*): as a shortcut, we shall refer to such a vector with the term "product". Naming with $G = \{g_i \mid i = 1,...,N\}$ the set of all available products for the consumption purpose *j*, it is exhaustively divided in *M* disjoint consumption segments G_h, with *h=1,...,M*. This partition (Γ_M) can be expressed as follows (to economize notation, hereafter we omit the suffix *j*):

$$\Gamma_M = \{G_h, h = 1,...,M\},$$

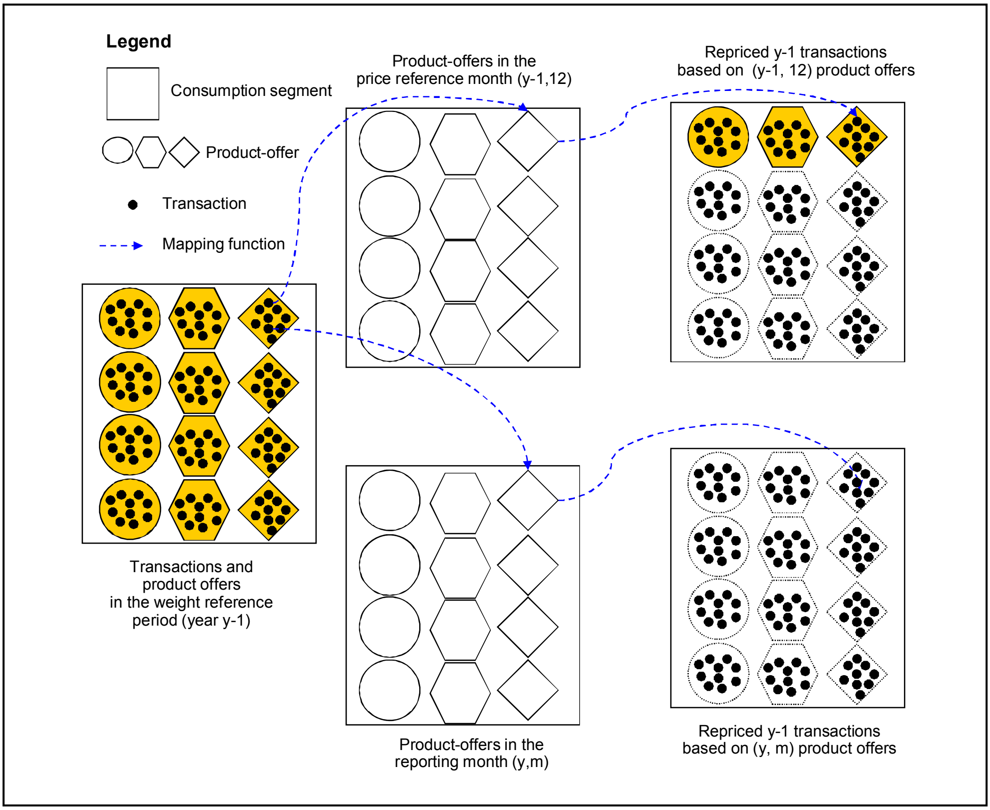
where $\bigcup_h G_h = G$ and $G_{h'} \bigcap G_{h''} = 0$ for every $h' \neq h''$. In order to simplify the definition of the universe and of the statistical target, some assumptions on the available product-offers are here introduced: the objective is to limit to price changes the possible sources of changes in the reference universe, and to provide a simplified framework for the definition of the statistical target (Dalèn 2001; Balk 2008, chap.5). The elements of the set *G* are assumed to be fixed and time-invariant: in other words, the number and the characteristics of the available offers do not change, and outlets and providers remain also unchanged (hypotheses A).

In this static environment, each vector g_i in the set *G* is associated to a second set of characteristics which describes the sequences of price spells and the corresponding time intervals of validity (p_i^t). From the definition of product-offer - recalled in par. 2.2 – each combination (g_i, p_i^t) describes a single product-offer. In order to control the number of product-offers associated to each product g_i we assume that discrete monthly pricing policies are adopted, where the prices of each element of *G* are eventually changed only at the very beginning of each month (hypotheses B).[10]

[10] This hypothesis may appear quite restrictive and not realistic, since monthly policies are quite rare. Nevertheless it is needed here in order to simplify notation and formalisation: the results can anyway be easily generalised, at least conceptually, to take account of intra-month policies.

Figure 2 describes the simplified framework derived from hypotheses A and B: it should be compared with the more general case reported in Figure 1. Product-offers can now be easily mapped given the invariance of G and the regularity of the sequence of price changes.

Figure 2 - A simplified framework for re-pricing



Given the hypotheses A and B, the generic element of the *(Nx13)* matrix Ω^{y-1} of all the product-offers in available year *y-1* is given by:

$$\left[g_i; \quad p_i^{y-1,1} \quad \dots \quad p_i^{y-1,12}\right].$$

The consumption expenditure (E) in the weight reference year *y-1* can be expressed as follows:

$$E^{y-1}\left(\Omega^{y-1}\right) = \sum_h \bar{p}_h^{y-1} T_h^{y-1} \tag{3}$$

where T labels the number of transactions and $\bar{p}_h^{y-1} = \sum_m p_h^{y-1,m}\left(\frac{T_h^{y-1,m}}{T_h^{y-1}}\right)$ is the annual average price actually paid for transactions T in the consumption segment h.

In order to define the "true" value of the target sub-index $I_j^{y,m}$ by adopting the consumption segments as the fixed objects defined in the HICP frame,[11] we need to simulate by means of re-pricing the total consumer expenditure of the weight reference year (*y-1*) on the basis of the product-offers available in the reporting month *m* of year *y* (identified by the couple (*y, m*)) and in the price reference month (*December y-1,* conventionally labelled with (*y,0*)). By applying (3) we obtain:

$$I^{y,m} = \frac{E^{y-1}\left(\Omega^{y,m}\right)}{E^{y-1}\left(\Omega^{y,0}\right)} = \frac{\sum_h \overline{p}_h^{y,m} T_h^{y-1}}{\sum_h \overline{p}_h^{y,0} T_h^{y-1}} = \sum_h I_h^{y,m} w_h \tag{4}$$

where $w_h^y = \frac{\overline{p}_h^{y,0} T_h^{y-1}}{\sum_h \overline{p}_h^{y,0} T_h^{y-1}}$ is the normalized value weight of consumption segment *h* and $I_h^{y,m} = \frac{\overline{p}_h^{y,m}}{\overline{p}_h^{y,0}}$ is its price relative.[12] Average prices are derived on the basis of a mapping of the set of product-offers Ω^{y-1} into the sets $\Omega^{y,0}$ and $\Omega^{y,m}$ available in the price reference and reporting months. The hypotheses A and B relating to set *G* make it possible to assume the existence of a one-to-one correspondence through mapping functions connecting product-offers: for each transaction involving product g_i in year *y-1*, the corresponding product-offers in the base and reference years are $(g_i, p_i^{y,0})$ and $(g_i, p_i^{y,m})$ respectively. Different versions of the target parameter defined in (4) can now be provided adopting alternative aggregation methods to calculate average prices. Two alternative approaches are proposed here, namely the weighted arithmetic mean:

$$\overline{p}_h^{y,m} = \frac{\sum_{g_i \in G_h} p_i^{y,m} T_i^{y-1,m}}{\sum_{g_i \in G_h} T_i^{y-1,m}} \tag{5}$$

and the geometric mean:

$$\overline{p}_h^{y,m} = \exp\left[\frac{\sum_{g_i \in G_h} T_i^{y-1,m} \ln\left(p_i^{y,m}\right)}{\sum_{g_i \in G_h} T_i^{y-1,m}}\right] \tag{6}$$

[11] See EC Regulation 1334/2007.

[12] Non zero average prices by segment in the base month (y,0) are here assumed; on the treatment of zero prices in the HICP see EUROSTAT (2001, p. 184-5).

Each approach implies specific assumptions on consumers' elasticity to price changes (see below, section 3.1). The aim, then, is to compare the properties of alternative standard approaches to sampling in order to provide an estimate for $I^{y,m}$.[13]

3. Sample size with alternative designs and aggregation formulas

3.1. Simple random sampling (SRS)

Assume that a simple random sample S of n products is drawn from G and to collect the prices of the corresponding product-offers in the price reference month and in a generic reporting month m. Given the hypotheses A and B, this is equivalent to drawing an identical sample of products from the sets of available product-offers $\Omega^{y,0}$ and $\Omega^{y,m}$. No other information is available on the universe of product-offers, consumption segments and transactions.

Different estimates of I^m can be produced[14] depending on the approach followed to aggregate the sampled quotes and to produce the target index estimates. Three alternative types of frequently used unweighted means are here compared: Carli (arithmetic mean of price relatives, labelled with "C"), Dutot (ratio of mean prices, "D") and Jevons (geometric mean, "J"), respectively:

$$\hat{I}^{m,C} = \frac{\sum_{i \in S} \frac{p_i^m}{p_i^0}}{n} = \frac{\sum_{i \in S} I_i^m}{n} \tag{7}$$

$$\hat{I}^{m,D} = \frac{\sum_{i \in S} p_i^m \Big/ n}{\sum_{i \in S} p_i^0 \Big/ n} = \frac{\sum_{i \in S} I_i^m p_i^0}{\sum_{i \in S} p_i^0} \tag{8}$$

$$\hat{I}^{m,J} = \exp\left[\frac{\sum_{i \in S} \ln\left(I_i^m\right)}{n}\right] \tag{9}$$

The relative convenience of aggregation formulas has been deeply debated in literature and it has been evaluated on the basis of their economic properties and on the characteristics of the underlying distributions of price changes. Each approach entails in fact specific assumptions on consumers' elasticity to price changes, which are reflected on

[13] This methodological framework is clearly open to a larger set of different approaches to aggregation. Even the stochastic approach can be considered, although it has been largely criticised, mainly for its weak economic foundations. Particular conditions concerning the distribution of price changes might anyway spur the adoption of this approach.

[14] Hereafter, we drop the suffix labelling the year.

the implicit weighting of transactions (Ilo 2004, chap.9; Leifer 2002, 2008; Viglino 2003; Balk 2003, 2008, chap. 5; Silver et al. 2006): Carli aggregation implies equal value weights for the product-offers and, for each product-offer, a constant expenditure in both the price reference and in the reporting month.[15] The Dutot formula implies equal and time-invariant quantities for each product-offer, whilst the Jevons formula assumes that the expenditure shares of the price reference month do not change when relative prices change so that some substitution due to the change in relative prices is therefore implied. With respect to the Carli formula, the Dutot approach assigns a higher weight to the product-offers with a higher price level in the price reference month and the Jevons approach assigns a higher weight to the product-offers with a lower price dynamics. Without going into the issue of the choice of the "right" formula, we want to discuss here some of their statistical properties in terms of precision within different sampling designs.[16]

Formula (7) provides an unbiased estimator of (4)-(5) only if the probability of selection is proportional to the weight of each product (Adelman 1958). The same applies to (9) with respect to the target set by expressions (4) and (6). For the Dutot formula (8) to be unbiased with respect to (4)-(5) it is necessary to add the condition that the price relatives be independent from the price levels in the price reference month (Balk 2008, chap.5).

Given a confidence level α and a relative error expressed as a share of the sample mean ($r = \beta I^{m,q}$, where $q \in \{C, D, J\}$), the adoption of these three methods implies some differences in the sample sizes needed to produce an error lower than the $\beta\%$ of the true value of the parameter with a probability of $\alpha\%$. These differences depend on the standard errors of the three types of sample means and on the form of their distributions. By adopting standard simple random sampling theory (Cochran 1977, chap. 4-6) separately for the three aggregation formulas, the necessary sample size in month *m* in the case of the Carli formula can be expressed as follows:

$$n_{SRS}^{m,C}(\alpha,\beta) \geq \left[\frac{t_\alpha}{\beta}\frac{s_{\hat{I}}^{m,C}}{\hat{I}^{m,C}}\right]^2 = \left[\frac{t_\alpha}{\beta}\right]^2 C_{I^{m,C}}^2 \tag{10}$$

where t is the corresponding value of the t-Student distribution, $s_{\hat{I}}^{m,C}$ is the standard error of the sample Carli mean and $C_{I^{m,C}}$ is the coefficient of variation of the Carli index (Cochran 1977, sect. 4.6; Ilo 2004, chap.5).[17]

For the Dutot formula we obtain:

$$n_{SRS}^{m,D}(\alpha,\beta) \geq \left[\frac{t_\alpha}{\beta}\right]^2 \left[C_{p^m}^2 + C_{p^0}^2 - 2\rho_{0,m} C_{p^0} C_{p^m}\right] \tag{11}$$

[15] Notice that the systematic use of the Carli formula has been banned for the estimates of the HICP (Commission Regulation (EC) No 1749/96, Art.7; EUROSTAT (2001, p.129, 155-156)).

[16] For discussions on this issue see for example Fenwick (2008) and Baskin et al. (1996).

[17] Hereafter, the sample fraction correction is not considered.

where $\rho_{0,m}$ is the Pearson's correlation coefficient of the price levels in the price reference and in the comparison month, while C_{p^m} and C_{p^0} are the coefficients of variation of the price series in the two months (Cochran 1977, sect. 6.3-6.5; Ilo 2004, chap.5).

In the case of the Jevons aggregation, the following expression is derived:

$$n_{SRS}^{m,J}(\alpha,\beta) \geq \left[\frac{t_\alpha}{2\log(x_\beta)}\right]^2 \left(s_{\log(I)}^m\right)^2 = \left[\frac{t_\alpha}{2\log(x_\beta)}\right]^2 C_{I^{m,J}}^2 \tag{12}$$

where $x_\beta = \frac{+\beta + \sqrt{\beta^2+4}}{2}$, $s_{\log(I)}^m$ is the standard deviation of the logarithms of the individual indices I_i and it is equal to the coefficient of variation of the Jevons mean $C_{I^{m,J}}$ (Cochran 1977, sect. 4.6; Ilo 2004, chap.5). The expression for x_β can be derived by applying the SRS formula for confidence interval to the log transformed variable and then transforming back and resolving by n. Following Norris (1940), $\hat{I}^{m,J} s_{\log(I)}^m$ corresponds to an estimate of the standard deviation of the geometric mean.

Expressions (10)-(12) derive sample size from the product between two elements: one dependent on α and β, and the other one is based on the coefficients of variation of indices and – in the case of Dutot aggregation - price levels. For reasonably low values of β (e.g., lower than 10%), the comparison among these formulas can be limited to this last element. In general, when the variability of prices and indices is very small, the three approaches lead to very similar sample sizes. On the contrary, some important differences might emerge when the variability indices and price levels is relatively large.

The Dutot index needs a higher sample size when there is a strong heterogeneity in price levels with negative or low positive correlation between price levels, and in particular when the largest price changes are associated with goods with a higher price level in the price reference month. In the case of the Jevons formula, the sample size tends to be relatively higher if the distribution of the price changes is negatively skewed while the opposite happens with a positive skewness.

3.2. Stratified random sampling (STRS)

It is reasonable to expect that a partition in consumption segments can potentially isolate homogeneous product-offers and, once adopted as a stratification criterion, may consequently reduce differences in the aggregation formulas. Nevertheless, in principle partitioning in consumption segments may not represent a good stratification criterion, and in any case this may not be the best way to control the variability of price changes. These two concepts are clearly conceptually distinct but may nearly coincide if the "economic" criteria adopted to define consumption segments meet also, as a by-product, the objective of isolating clusters of products characterised by homogeneous pricing policies. Consumption segments should be based - according to the HICP legal based recalled in section 2.2 – on supply and demand side market analysis, and it is very likely then that they can target well the variability of price changes. The definition of the border between segmentation and stratification very much depends on the ambiguity – reminded also in HICP sampling regulation - on the level of aggregation, that is on how deep the segmentation is run. Such an issue deserves deeper case studies necessarily based also on survey microdata. Here we implicitly chose to collapse the two concepts and to compare the performance of different degrees of aggregation, from no stratification (segmentation) at all to deeper stratification.

We assume that more information is available concerning the consumption expenditure in the weight reference year: the true weighting structure (w_h) of a partition in consumption segments (Γ_M) is known, although no other information is available within each segment concerning the expenditure shares of the product-offers. It is important to notice here that weights are not identified here as a potential source of errors: this practice is common to most of the approaches to the measurement of the statistical error in CPI estimates (Biggeri et al. 1987 is a meaningful exception). In this work we adopt this same hypothesis, although we are perfectly conscious that additional work needs to be done in order to join this analysis of price and price indices variability with that of the precision of weighting: the latter is of paramount importance in order to evaluate the effectiveness of deeper stratifications.

If a stratified random design is adopted, the estimate may be obtained as a value-weighted arithmetic mean of the indices of each segment (stratum):

$$\hat{I}_{STRS}^{m,q} = \sum \hat{I}_h^{m,q} w_h \ .$$

The standard deviations within each stratum, for the three alternative formulas, will be given by:

$$s_h^{m,C} = \frac{\sum_{i \in h}\left(I_i - \hat{I}_h^{m,C}\right)^2}{n_h}$$

$$s_h^{m,D} = \frac{1}{n_h\left(\bar{p}_h^0\right)^2}\left[s_{p_h^m}^2 + \left(\hat{I}_h^{m,D} s_{p_h^0}\right)^2 - 2\hat{I}_h^{m,D} s_{p_h^0 p_h^m}\right]$$

$$s_h^{m,J} = \hat{I}_h^{m,J} s_{\log(I_h)}^m$$

Independently of the type of elementary aggregation, total sample size with optimal allocation can be expressed as follows:

$$n_{STRS}^{m,q}(\alpha,\beta) \geq \left[\frac{t_\alpha}{\beta}\right]^2 \left[\frac{\sum_h s_h^{m,q} w_h}{\sum_h \hat{I}_h^{m,q} w_h}\right]^2 \qquad (13)$$

where $s_h^{m,C}$ is the standard deviation within stratum h (Cochran 1977, sect. 5.4-5.9).[18]

[18] As in the case of SRS, the sampling fraction correction has been skipped.

Following the optimal allocation per strata (i.e. proportional to the standard deviation), sample size in each stratum can be expressed as follows:

$$n_{STRS,h}^{m,q}(\alpha,\beta) = n_{STRS}^{m,q}(\alpha,\beta)\frac{s_h^{m,q} w_h}{\sum_h s_h^{m,q} w_h} \tag{14}$$

Expression (13) suggests that stratified designs can reduce the source of discrepancies among aggregation formulas, depending on the ability of the former to reduce the variance within strata by means of a clustering approach able to isolate the criteria used to define pricing policies.

3.3. Alternative temporal targets

In the simplified framework under hypotheses A and B, expressions (10)-(12) and (13)-(14) have been for the moment referred to a generic reporting month *m*. They fix the number of product-offers whose price must be collected in the price reference month (*December y-1*) and in the generic reporting month *m* in order to achieve the desired precision level for the estimate of the price index in *m*. Nevertheless, if our aim is to produce complete annual series of monthly estimates, then a number of consequences do emerge, depending on the way we approach this task. The sample size needed for the estimates referred to month *m* is in fact in general different from the one needed to arrange the same precision for another month *m'*. This happens because the nature of price dynamics possibly changes from month to month in a way which may depend on the specific demand and supply characteristics of each consumer market.[19] Monthly samples can differ substantially, especially in the case of seasonal goods or services.

In any case the price collection in *December y-1* provides the base for the annual link, and therefore its role is crucial for all the monthly estimates that we are targeting: if we target a minimum precision level in every month, the sample in this price reference month has to be drawn in order to satisfy the size requirements of all the twelve following months. In particular, in SRS designs, sample size in the price reference month must be equal to the maximum size needed in the twelve months:

$$n_{SRS}(monthly) = \max_m (n_{SRS}^m(\alpha,\beta)) \tag{15}$$

If on one side the price collection in *December y-1* is the largest one, it might be not necessary to activate a monthly price collection extended to all this sample for the entire sequence of twelve monthly estimates. It is in fact possible to modulate price collection according to the actual monthly needs based on expressions (10)-(12). If we know that in a given month the expected variability is very low and that the desired precision can be achieved with a sample which is half the one drawn for the base according to (15), than we

[19] We may have to do with a very heterogeneous set of consumer markets - seasonal products, highly competitive markets, oligopolistic or monopolistic markets, markets highly dependent on external influences (markets for international commodities, weather, natural events) or even administered prices, and so on -, all behaving in quite different ways and with a variety of pricing policies. For a classification of price index dynamics within the HICP see De Gregorio (2011).

can save resources for price collection and concentrate them, for example, for the most critical months. A modular approach to price collection is then possible, and it gives the possibility to change the sample size every month in order to assure a given precision target in presence of heterogeneous variability patterns observed across months. In the case of seasonal products, for instance, this approach requires the largest effort in price collection in the price reference month, a high activation rate of the sample during peak months and lower off-peak rates.[20]

With stratified designs some further complications may arise since allocation is also a relevant factor. The sample size in the price reference month derives, in fact, from the sum of the largest monthly size of each stratum:

$$n_{STRS}(monthly) = \sum_{h} \max_{m}(n_{STRS,h}^{m}(\alpha, \beta)) \tag{16}$$

which may be much larger than the maximum overall monthly size derived from expression (14). This happens, in particular, whenever peaks in variability have distinct time patterns across strata, such as in markets characterized by seasonal pricing where peak months generally show higher variability: the timing of seasonal peaks, in fact, might differ across strata and this mere fact induces the need of larger samples in the price reference month. Something similar might happen in sectors characterized by highly irregular patterns.

The sub-indices with a relatively large variability or those characterized by seasonal behaviours are indeed only a part of the whole set of HICP sub-indices. It has been estimated that within the euro zone between 2004 and 2008 about 25% of HICP four-digit sub-indices showed a relatively strong monthly dynamics whilst about 7.3% showed a clear seasonal pattern (De Gregorio 2011). For what concerns the remaining indices, they were referred to markets where, price changes were quite regular and very slow, at least in periods of low inflation. In such cases, in the first months of the year – which are nearer to the base of *December y-1* - most observations are concentrated in the “no-change” zone: as a consequence, the distribution of price changes in those months is positively skewed. This asymmetry progressively loses ground as one moves away from the price reference month towards the final part of the year. If on one side the inertia of price indices in the first months reveals a very low variability and hence lower sample size needs, on the other side it might generate complications since the hypothesis of normality could not apply.

In any case, due to inertia, the last months of the year might be those in need of the largest samples, and the adoption of the annual link of December as a primary target for the estimates appears extremely reasonable: its importance relies in fact on the permanent effect that the link has on the chained index *H*.[21] In the case of the two types of design discussed above we obtain:

$$n(link) = n^{12}(\alpha, \beta) \tag{17}$$

This formula bears relevant gains in sample size with respect to expressions (15) and (16) only if the variability of price changes is diluted during the year and it is not

[20] For an application of this modular approach to seasonal products see De Gregorio, Munzi et al. (2008).

[21] See expression (1); Fenwick (1999) examines the issue of the choice of the price reference month, emphasizing the problems that may arise in the choice of the aggregation formula in case of large variability of price dynamics.

concentrated in the final month. Alternative reasonable targets might be set on quarterly or yearly averages:

$$n(quarterly) = \max_Q (n^Q(\alpha, \beta)) \tag{19}$$

and

$$n(yearly) = n^Y(\alpha, \beta) \tag{20}$$

In particular, for quarterly targets in stratified designs, allocation effects must also be considered as in the case of monthly estimates. It is also possible to use combined targets, for instance to guarantee the precision level on quarterly and annual link estimates.

4. Two case studies

4.1. Artificial populations

In order to test the combined effect of sample design and aggregation formulas on the variance of the estimates, we generated two artificial target populations starting from a selection of the microdata collected by ISTAT for the 2007 cycle of the HICP, and we iterated the extraction of samples from these populations in order to estimate the target parameter defined in Section 2. In particular, two case studies are here presented.[22] They are referred to price series characterized by high variability and heterogeneous behaviours: the first case regards European air transports, where the high volatility of price changes is partly explained by seasonal patterns; the other one regards package holidays, strongly affected by overlapping seasonal peaks with some inertia in the first months of the year.[23]

More formally, following the simplified approach outlined in par. 2.3, each set of microdata is interpreted as if it was a random sample drawn from the product-offers available in year y ($Z \subset \Omega^y$). Each record is characterized by a product identifier (g_i) and by a vector of 13 price quotes - from month *0* (the price reference month, namely December 2006) to month *12* (December 2007). For each market, a detailed and exhaustive partition of the goods in M_0 disjoint sets of consumption segments is then given:

$$\Gamma_{M_0} = \{G_h, h = 1, \dots, M_0\}.$$

[22] Official microdata have been treated here with a different purpose from that pursued by ISTAT; it follows that results cannot be compared at any rate with the official figures currently disseminated.

[23] Flights and package holidays are both identified in De Gregorio (2011) as the sub-indices with the most heterogeneous behaviours across the countries of the euro zone, possibly needing further harmonization. For a methodological overview of the methods actually adopted by ISTAT to estimate these indices, see ISTAT (2009) and De Gregorio, Fatello et al. (2008).

Alternative but less detailed partitions Γ_{M_i} might be obtained by hierarchical aggregation of the subsets of Γ_{M_0}. For each partition a vector of normalized weights is accordingly defined:

$$W_{M_i} = \left\{ w_h, h = 1,..., M_i \mid \sum_h w_h = 1 \right\}.$$

Microdata in each set Z actually derive from stratified samples which have not been selected with probabilistic rules (ISTAT 2009; De Gregorio, Fatello et al. 2008, p. 20, 28-32). Nevertheless, they are treated here as if they were derived from random selections and the element of each set are expanded proportionally to the weighting structure W_{M_0} in order to form an infinite population. K simple random samples, each of n product-offers, are finally drawn from these infinite artificial populations. The yearly series of the monthly estimates $\hat{I}_k^{m,q}$ ($k=1,...,K$) are derived from each sample, adopting alternatively the Jevons, Dutot or Carli aggregation (expressions (7)-(9)). An inductive estimation of the sample mean variance is then produced and, consequently, an estimate of the sample size by means of the formulas derived in the preceding sections is provided. An identical approach is used to estimate the sample size for stratified designs based on alternative partitions of the target population.

All the simulations for the markets under scrutiny have been made by extracting iteratively 300 samples of 500 products each. Given a 1% error and a 95% confidence level, distinct temporal targets have been separately considered. Tables 1 and 2 (see par. 4.4 below) describe a relative measure of the sample size calculated as a multiple of a benchmark size (the one needed to estimate the yearly average with Carli aggregation and SRS). In particular the sample sizes have been determined in order to obtain the desired precision level for alternative temporal targets: i.e. separately for each single month, the cumulative target extended to the whole set of months (adopting expression (15) and (16)), the quarterly and yearly averages (expressions (18) and (19)). The desired precision target has been finally set on the link month of December, which - given the chaining procedure - affects permanently the fixed base series (expression (17)).

For the construction of the artificial population, in the case of European air transports we have used data from the original sample of $N=328$ product-offers, concerning as many European return flights connecting the country of origin (national) with the other countries (foreign). Each return flight is defined by a national and a foreign airport area (for instance, Rome and Frankfurt).

Four distinct partitions are used to provide alternative exhaustive segmentations of the target population. An elementary stratification Γ_{51} (51 strata) establishes an exhaustive segmentation by national and foreign regions (sub-national) and by type of carrier (low cost vs. full service carriers). A less detailed partition collapses the regions of each foreign country (Γ_{38}); a further aggregation of consumption segments uses only the country of destination and the type of carrier (Γ_{15}), and the less detailed partition only the country of destination (Γ_{11}). An elementary consumption segment can identify, for example, the low cost flights from the region A1 in country A (national) to the region B1 in the foreign country B; less detailed partitions identify, orderly, all the low cost flights from region A1 to B, all the low cost flights from A to B and all the flights from A to B.

In the case of package holidays[24] we have used the data from an original monthly sample of *N=246* records. Each package is defined by a region of destination. Two distinct partitions provide exhaustive segmentations of the target population. A more detailed stratification Γ_{43} splits the universe into countries and type of holiday (i.e.: sea, mountain, city, etc.), while a less detailed segmentation Γ_{12} adopts only the splitting by country. As an example, an elementary segment could be the market for package holidays for type of holiday A1 in country A; a less detailed partition would concern all the packages for holidays for country A.

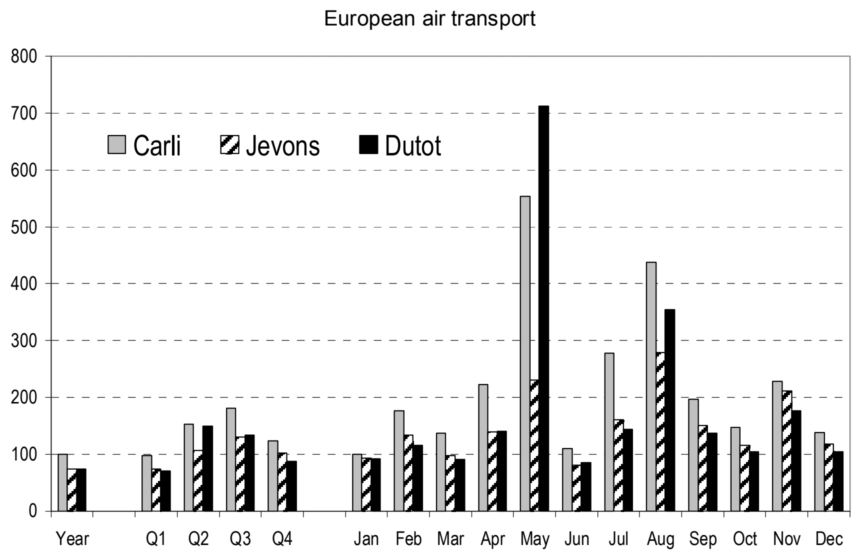
4.2. Design effect for independent temporal targets

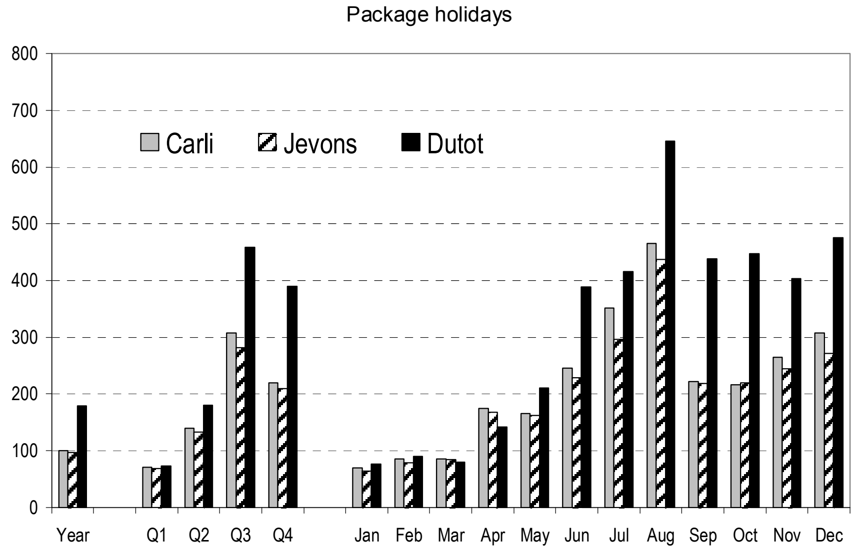
Chart 1 plots the sample sizes needed with SRS for each month and quarter, and for the yearly average (in all the charts and tables, the sample size needed to estimate the yearly average with SRS and Carli aggregation is used as a benchmark and has been set equal to 100). Both markets show large differences in the variability within each month; Jevons formula delivers the best performance and Carli the worst, although heterogeneity in price levels seriously impairs the performance of Dutot aggregation during seasonal peaks; quarterly and yearly targets are far less demanding, although inertia effects may require larger samples in the last quarters.

In particular, for the separate estimates of the monthly indices of air transports smaller samples are needed at the beginning and at the end of the year (and in June) whilst the largest sizes are found in association with seasonal peaks in May and August (several times higher then the benchmark): in these months, in fact, the distributions of both price levels and price changes are positively skewed. Jevons aggregation is relatively less demanding, since it requires in May a sample size laying between two and three times the benchmark; in the same month the Dutot formula delivers by far the worst result (seven times the benchmark). Carli aggregation needs the largest size in eleven months out of twelve (with the median monthly size more than 35% higher than Jevons'). Dutot generates lower sample sizes in most of the off-peaks months (first and fourth quarter) due to a more appreciable homogeneity in price levels.

[24] Only foreign travels were considered.

Chart 1 - Sample size with SRS by sub-index, temporal target and type of aggregation (Indices. Base: size for yearly target with SRS and Carli aggregation = 100)





Moreover, although in air transports strong seasonal fluctuations hide any effect related to the time-distance between the reporting and the reference months, in the case of package holidays seasonal and inertia effects are combined: after the summer peaks (in July and August) sample size remains in fact quite large as compared to the first months of the year. This is the effect of the inertia of price dynamics, since price levels in the first months tend to range closer to their reference level and the estimates are less challenging. The annual link, in particular, seems to need a large sample as opposed to air transports where the link month was one of the easiest targets. In package holidays the worst performance is provided by Dutot aggregation: it works relatively well at the beginning of the year, but soon becomes by far the less appropriate (in terms of sample size) in the remaining months. This is due, probably, to the high heterogeneity of price levels, since they vary considerably across markets and show some likely correlation with price changes.

The figures for quarterly targets partially confirm this picture, although they are quite smoother for air transports where sample size never doubles the benchmark: Q1 requires the same sample as January or the yearly average separately for each aggregation method; Q2 and Q3 are more demanding, although they never double the sample size needed to estimate the yearly average. Package holidays on the contrary demand larger efforts in the last two quarters, due to the inertia effects, and confirm the inadequacy of Dutot aggregation, while Carli and Jevons require nearly the same sample size for all the quarters and for the yearly average.

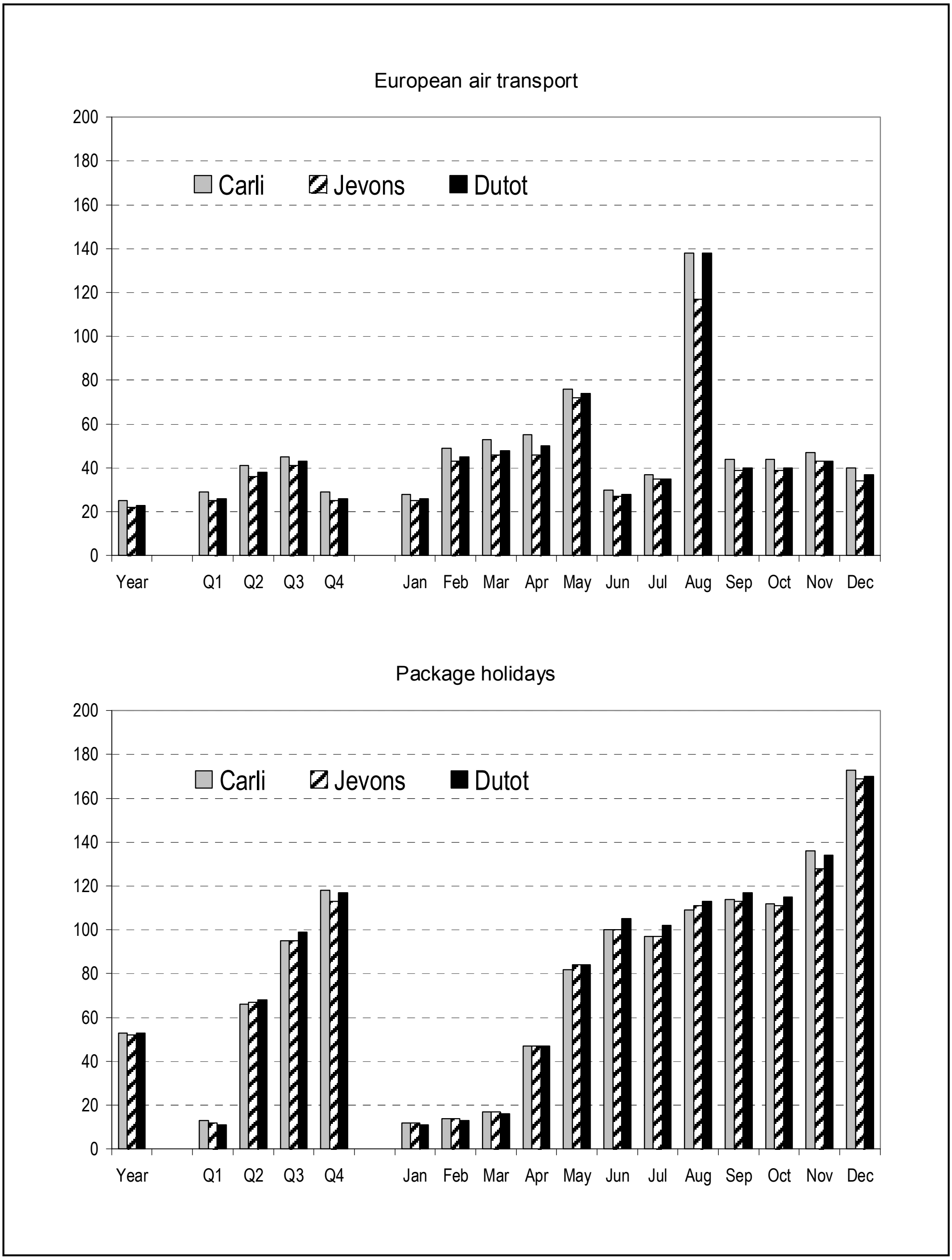
Chart 2 reports the effects on sample size deriving from the adoption of STRS at the most detailed level of stratification.[25] The effects of stratification are quite impressive: sample size is strongly reduced, the seasonal peaks are considerably smoothed and the differences among aggregation formulae tend to disappear.

In air transports the sample size necessary to meet the yearly target is slightly more than 20% higher than the benchmark; quarterly samples and monthly samples are strongly reduced to 25-30% of the corresponding need in a SRS frame. Such decrease is particularly strong in peak months, especially in May. The effect of stratification is stronger with the Carli formula where the performance in terms of sample size improves considerably (sample size is only 10% higher than Jevons, in median). With the introduction of stratification, Jevons aggregation is still the one systematically requiring smaller samples: nevertheless, with stratification the differences in sample size due to alternative approaches to aggregation tend to shrink considerably.

This particular aspect is also evident in package holidays; stratification removes seasonal effects and only inertia plays a major. For the first three months samples are less the 20% of the benchmark, in June they pass 100% and the link month is the more demanding (nearly 170%). Quarters behave similarly, and this effect plays a key role in sustaining also the size of the sample required to meet the yearly target.

[25] Please, notice the different scale of this chart as compared to Chart 1.

Chart 2 - Sample size with STRS by sub-index, temporal target and type of aggregation (Indices. Base: size for yearly target with SRS and Carli aggregation = 100)



4.3. Combined temporal targets and modular price collection

In the previous sections, we discussed monthly or quarterly targets where each time span was considered independently from any other. What is more interesting is to see how large the sample has to be in order to meet at the same time all the monthly targets or all the quarterly targets. For SRS the solution is trivial: it is in fact sufficient to adopt the maximum monthly or quarterly sample sizes. On the contrary, with stratified designs strata allocation effects might complicate the matter (see par. 3.3): the relative efficiency of Jevons aggregation looses part of its advantage as compared to Dutot and Carli when combined targets are pursued, since Jevons allocation tends to show a higher heterogeneity in the sample size needed each month in each stratum. It can be said that the adoption of combined targets and stratified designs brings towards a reduction in the differences in efficiency due to the aggregation method.

More specifically, for the whole set of monthly targets of air transports with SRS the use of a Carli aggregation would need nearly 5.54 times the benchmark (Table 1). The Dutot formula delivers an even worse result (7.12), due to the high heterogeneity in price levels. The Jevons approach (2.79) needs half the sample size as compared with Carli. Such large samples derive from the high volatility observed for price levels and indices in peak months. If we reduce the SRS target to quarterly estimates, the sample sizes shrink drastically (between 1.30 to 1.81 times the benchmark) and the differences among the methods also are strongly reduced. The yearly estimates need nearly half the sample used for the quarterly target, while the annual link of December is placed between the quarterly and the yearly target. As long as the yearly target is concerned, Dutot equals Jevons' performance.

The effects of stratification are confirmed for combined targets: sharp reduction in sample size and more homogeneous results across the three aggregation approaches. The introduction of the first two levels of stratification brings large improvements, in particular for the Carli formula. The partition in 38 strata is extremely fruitful for all types of formulae, while the most detailed partition brings a comparatively minor reduction in sample size. In the passage from SRS to the most detailed stratified design there takes place a reduction of almost 70% of the necessary sample size. The temporal patterns of variability within strata are quite differentiated across months: consequently, the allocation effect induces appreciable differences between the sample size needed to target the whole set of monthly prices and the maximum size for separate monthly targets. If we consider the whole set of monthly targets (see Chart 1 and formula (16)), the sample size for Carli and Dutot aggregation is nearly 20% higher than the maximum size shown in Chart 1; Jevons formula, although it is in general more efficient, needs a sample nearly 30% higher than the respective maximum (151 vs. 117). For the quarterly indices, the size increase needed to meet all the monthly targets is slightly above 20%. The yearly target, independently of the design, requires about 15-20% of the sample size needed for the monthly targets, and the link demands nearly 30%.

In general Jevons aggregation performs better, with some exceptions where Dutot appears less demanding. Carli generally implies larger samples, although the differences collapse as stratification runs deeper. The heterogeneity of price levels damages the performance of the Dutot formula, especially where stratification is absent or limited, while the sample size derived from the Jevons formula appears less influenced by the presence of larger prices. If we consider package holidays, the irregularity of the monthly variability within strata appears less pronounced as compared to air transports (Chart 1). The SRS sample size needed to target the whole set of twelve months is in fact only 5% higher than the maximum size needed to meet separately the monthly targets (Table 1): Carli and Jevons aggregations require similar sample sizes (respectively, 4.65 and 4.37 times the benchmark), while Dutot has the worst performance. Most of the gains are obtained with the first level of stratification (12 strata): the three methods are almost equivalent. The adoption of further stratification confirms this picture and strongly reduces size needs. The equivalence of aggregation methods induced by stratification appears even stronger then in the case of air transports. It is worth the while to notice that, in any case (and differently from air transports), the estimate of the annual link for package holidays is more demanding than the estimate of quarterly and yearly targets, and that as stratification is adopted the estimate of the annual link requires a sample size which very near to the one needed to estimate the whole set of monthly targets.

Table 1 - Sample size, by sub-index, temporal target, type of aggregation and sample design (Indices. Base: size for yearly target with SRS and Carli aggregation = 100)

DESIGN	Aggregation	Temporal target			
		Monthly	Quarterly	Yearly	Annual link
		EUROPEAN AIR TRANSPORT			
SRS	Carli	554	181	100	138
	Jevons	279	130	74	118
	Dutot	712	149	74	104
STRS 11	Carli	329	149	81	105
	Jevons	239	105	59	81
	Dutot	284	116	63	88
STRS 15	Carli	280	106	61	90
	Jevons	225	90	50	74
	Dutot	267	98	55	81
STRS 38	Carli	191	65	32	55
	Jevons	171	59	28	48
	Dutot	188	61	29	50
STRS 51	Carli	169	55	25	40
	Jevons	151	49	22	34
	Dutot	167	52	23	37
		PACKAGE HOLIDAYS			
SRS	Carli	465	307	100	307
	Jevons	437	282	97	271
	Dutot	646	458	179	476
STRS 11	Carli	258	183	78	245
	Jevons	257	182	77	234
	Dutot	254	179	77	241
STRS 43	Carli	183	123	53	173
	Jevons	184	123	52	169
	Dutot	182	123	53	170

The pursuit of monthly targets might be very expensive and very much influenced by a few peak months. For this reason it might be redundant to extend price collection to the whole sample every month. It is instead possible to adopt cost-effective solutions based on modular approaches to price collection: only part of the sample might be surveyed every month, depending on the size needs imposed by the peculiar characteristics of variability in that month. The adoption of a modular month-dependent size to meet monthly or quarterly targets implies in any case a larger sample in *December y-1* – whose size can be derived from Table 1 - and reduced price collection especially in the months or quarters where the variability is lower.[26]

Table 2 reports some evidence. As we have seen, in the case of European air transports with SRS and monthly targets, the size index of the sample needed in the price reference month would be 554 in the case of Carli aggregation. Anyway, in the overall period of 13 months to December *y,* price collection can be skipped for more than a half (54.4%) of that sample. The saving is even higher in the case of Dutot aggregation (68%), while Jevons aggregation brings to a relatively lower saving (42.3%). Something similar happens for air transport when SRS is applied to quarterly targets, although in this case the sample sizes in the price reference month are much more homogeneous: average saving is around 20%, slightly for Dutot and Jevons. Stratification, as we have seen before, reduces the differences among aggregation methods and also the savings in price collection are quite similar: two units of the base sample out of three are saved on average with monthly targets and one out of three with quarterly targets.

Table 2 - Sample size reduction for price collection with modular sampling, by design, temporal target and type of aggregation (Indices. Base: size for yearly target with SRS and Carli aggregation = 100)

TEMPORAL TARGET PRICE COLLECTION	SRS			STRS (a)		
	Carli	Jevons	Dutot	Carli	Jevons	Dutot
			EUROPEAN AIR TRANSPORT			
Monthly target						
December y-1	554	279	712	169	151	167
Average 13 ..	252	161	228	62	55	59
Saving (%)	54,4	42,3	68,0	63,2	63,4	64,5
Quarterly target						
December y-1	181	130	149	55	49	52
Average 13 ..	142	105	113	37	33	35
Saving (%)	21,7	19,0	23,8	31,8	33,0	33,3
			PACKAGE HOLIDAYS			
Monthly target						
December y-1	465	437	646	183	184	182
Average 13 ..	300	279	453	123	122	125
Saving (%)	35,4	36,1	30,0	32,8	33,8	31,5
Quarterly target						
December y-1	307	282	458	123	123	123
Average 13 ..	194	181	289	77	76	78
Saving (%)	36,9	35,7	36,8	37,8	38,4	37,0

(a) Only the most detailed stratifications are considered here, i.e. 51 strata for European air transport and 43 strata for package holidays.

[26] Modular sample sizes are forcedly adopted in some specific markets, like in the case of accommodations in sites characterised by a strong seasonality (De Gregorio, Munzi et al. 2008).

Differently, in the case of package holidays with a SRS design Dutot aggregation is not only the less efficient method but also the one with the lowest saving deriving from the modular approach. With both SRS and STRS designs, slightly larger savings are obtained for the estimates of quarterly targets (between 35% and 40%).

Concluding remarks

This work has investigated and empirically tested some aspects of sample designs derived from the application of the most recent advancements in HICP methodology, by assuming a static definition of the target consumer market where replacements and changes in the range of products are excluded (see par. 2). As a whole, HICP concepts and methodology appear very well suitable for a more explicit use of the concepts and tools of statistical inference to estimate consumer price indices and to evaluate the quality of the estimates: for this reason, they also pave the way for a more cost-effective planning the technical management of monthly surveys.

In particular, we have analysed and compared the sample sizes requirements by combining the adoption of simple or stratified random sampling with alternative approaches to elementary aggregation and with a set of temporal targets. As a first step, we derived in par. 3 the expressions for a generic monthly sample size by type of design and aggregation, and developed them as functions of the coefficient of variation of indices and price levels: our findings confirm that aggregation effects on optimal sample size depend crucially on the level of relative variability and skewness of observations; such effects tend to annul when price changes are smoother and if the precision target is sufficiently tight.

The case studies reported in section 4 confirm these results and highlight some more points: the crucial role of stratification in saving sample size; the heterogeneity of the results obtained with different approaches to aggregation, and its fading out as stratification is introduced and when allocation effects are at work in stratified designs with multiple temporal targets; the possibility to adopt modular schemes of price collection especially with strongly seasonal items; the options opened by fixing temporal targets alternative to the monthly series, especially quarterly averages or the annual link; the role of indices' inertia in the determination of the sample size of the annual link.

Empirical evidence shows that stratification may shrink sample size by 50% to 70% as compared to SRS design. The choice of the strata is of paramount importance, since it involves theoretical and microeconomic issues: here it has been based on marketing criteria, trying to isolate possibly homogeneous consumption segments and clusters of pricing policies. The issue of how deep stratification should be is also very important. The introduction of a first layer with a few strata brings immediately large gains in sample size. More complex stratifications usually - but not necessarily - produce comparable gains with respect to more elementary designs. This depends obviously on the relative efficiency of a deeper stratification to compress the variance within strata. In the case of air transports, for example, adding the type of carrier to the country of destination increases by nearly 40% the number of strata but does not seem to generate very large gains, at least with the Jevons or Dutot aggregation. On the contrary, more detailed areas of destination produce important size gains, since they probably better reflect the pricing criteria of this market.

The case studies all emphasise the role that market segmentation and stratification have in reducing optimal size, increasing precision and saving resources. Stratified designs can produce other interesting effects, such as reducing drastically the heterogeneity resulting from alternative aggregation methods. Very heterogeneous optimal sizes might in fact derive from Carli, Dutot or Jevons approaches, depending on the variability of price changes and price levels. It is well known that with no - or just with a few - strata, Jevons performs significantly better in terms of optimal size. Deeper stratifications tend anyway to reduce this advantage due to within-strata homogeneity. The choice of the aggregation method is an issue largely debated in literature but it loses importance as stratification is considered, especially with large and highly stratified samples. Even the adoption of combined targets produces some smoothing for the aggregation effect: empirical evidence suggests that Jevons aggregation loses part of its advantages when the target is moved from a single month to the whole series of monthly indices, due to a less favourable monthly allocation of the units across strata.

The fact that optimal sample size might be determined on a monthly basis has also a number of consequences. Even if we stick to a defined approach to sampling and aggregation, heterogeneous monthly results might occur in markets where the variability of pricing behaviours is monthly dependent. This is likely to happen with seasonal items or even in markets where the variability of prices is somewhat structural: the cases of flights and package holidays are paradigmatic. In such a context, adopting a constant monthly size in price collection appears sub-optimal. What we intend to highlight is that a modular approach to price collection is possible, allowing a concentration of resources in those months where variability hits a peak, and consequently favouring a better management and scheduling of the surveys. Empirical evidence suggests the adoption of a modular price collection, with strong efforts concentrated in the price reference month while part of the monthly samples can be drastically reduced.

In this respect, the consideration of alternative temporal targets also appears as a strategic issue, if one of the objectives is to save resources by optimising their use. When quarterly targets are concerned, considerable gains in sample size are obtained as compared to monthly targets: large differences among aggregation methods anyway persist also if the target is moved on the yearly average or on the annual link and unless highly stratified designs are considered. Targeting the annual link is justified by its permanent effect on the chained index: such target may imply a large gain in sample size, as it happens for air transports; but if the link month is among those showing a higher variability (as in the case of package holidays) this objective may not produce large enough gains.

Although this work is based on several restrictive hypotheses on the dynamics of the set of the available product-offers (time invariance, with no changes in the range of the products and in the retail network), such hypotheses were essential in order to provide a reliable definition of the statistical target and a one-to-one mapping for the re-pricing of the set of the transactions in the weight reference period based on the product-offers available in the price reference and in the reporting month. This can be interpreted as a first approximation: relaxing these hypotheses implies in fact a huge modelling of consumers' choices in order to produce more sophisticated mapping functions. Further developments on these issues might be obtained both on the theoretical and empirical grounds. Concerning the first, the pioneering work of Ribe (2000) deserves more analysis on the form and nature of the mapping functions and their implications, especially with reference to the structural characteristics of consumer markets. It could be fruitful to consider different classes of mapping functions to be used in particular clusters of consumer markets. Empirical studies might help in this work, by examining other sectors and by providing deeper insights on the relative efficiency of alternative stratification criteria. Under this respect, the ambiguity of consumption segments and the role of stratification need also further empirical research, especially for what concerns the study of supply and demand effects on specific consumer markets on a case-by-case basis in order to isolate the sources of pricing behaviour.

A further remark concerns the weighting strategy. In this paper it was assumed that weights are not a source of potential statistical error, although weighting are estimates themselves and are a primary source of error, being often at the core of the criticism against official CPI estimates. Nevertheless, a specific and structured literature on the subject is lacking (remarkable exceptions, such as Biggeri et al. 1987, do not impair this statement), although the adoption of confidence intervals and precision targets cannot ignore this issue. Work on this subject is thus necessary, with the objective to join together the effects of price and price indices variability with those of weights variability. Finally, the role of overall inflation has also to be considered: expected variability of price levels and price changes is strictly connected with the expected evolution of inflation expectations. This aspect also should be modelled in order to achieve a more complete approach to CPI sampling: quite surprisingly, also in this case literature is lacking.

References

Adelman, I. 1958. A new approach to the construction of index numbers. *Review of economics and statistics*, 40, p. 240-249.

Andersson C., G. Forsman, J. Wretman. 1987. On the measurement of errors in the Swedish consumer price index. Invited paper 12.2 at the 46th session of the ISI-IASS meeting, Tokio 8-16 September. Booklet, p.261-277.

Balk, B.M. 2003. Price indexes for elementary aggregates: the sampling approach. Paper presented at the seventh meeting of the International working group on price indices, Paris, 27-29 May.

Balk, B.M. 2008. *Price and quantity index numbers. Models for measuring aggregate change and difference*. New York: Cambridge university press, p.283.

Banerjee, K.S. 1956. A note on the optimal allocation of consumption items in the construction of a cost of living index. *Econometrica*, Vol. 24, No. 3 (July), p. 294-295.

Baskin, R.M., S.G. Leaver. 1996. Estimating the sampling variance for alternative forms of the U.S. consumer price index. *Proceedings of the Survey research methods section, American statistical association*, p. 192-197.

Biggeri, L., P.D. Falorsi. 2006. A probability sample strategy for improving the quality of the consumer price index survey using the information of the business register. Paper presented at the joint ECE/ILO meeting of the of Group of experts on consumer price indices, Working paper n.12, Geneva 10-12 May.

Biggeri, L., A. Giommi. 1987. On the accuracy and precision of the consumer price indices. Methods and applications to evaluate the influence of the sampling of households. Invited paper 12.1 at the 46th session of the joint ISI-IASS meeting, Tokio 8-16 September. Booklet, p.244-260.

Cochran, W.G. 1977. *Sampling techniques*. 3rd edition. New York: John Wiley and sons, p. 428.

Dalèn, J. 1992. Computing elementary aggregates in the Swedish consumer price index. *Journal of official statistics*, Vol. 8, No.2, p. 129-147.

Dalèn, J. 1998. On the statistical objective of a Laspeyres' price index. Paper presented at the fourth meeting of the International working group on price indices, Washington D.C., 22-24 April.

Dalèn, J. 2001. Statistical targets for price indexes in dynamic universes. Paper presented at the sixth meeting of the International working group on price indices, Canberra, Australia, 2-6 April.

Dalèn, J., E. Ohlsson. 1995. Variance estimation in the Swedish consumer price index. *Journal of business & economic statistics,* Vol.13, No. 3 (July).

De Gregorio, C. 2011. The variability of HICP sub-indices and harmonisation needs. *Rivista di statistica ufficiale*, n.1 p. 5-32.

De Gregorio, C., S. Fatello, R. Lo Conte, S. Mosca, F. Rossetti. 2008. Sampling design and treatment of products in ISTAT centralised CPI surveys. *Contributi ISTAT,* n. 1.

De Gregorio, C., C. Munzi, P. Zavagnini. 2008. Problemi di stima, effetti stagionali e politiche di prezzo in alcuni servizi di alloggio complementari: alcune evidenze dalle rilevazioni centralizzate dei prezzi al consumo. *Contributi ISTAT*, n. 6.

De Haan, J., E. Opperdoes, C. Schut. 1997. Item sampling in the consumer price index: a case study using scanner data. Paper presented at the joint ECE/ILO meeting of the Group of experts on consumer price indices, working paper n.1, Geneva 24-27 November.

Dippo, C.S., K.M Wolter. 1983. A comparison of variance estimators using the Taylor series approximation. Proceedings of the Survey research methods section, American statistical association, p. 113-119.

Dorfman, A.H., J. Lent, S.G. Leaver, E. Wegman. 2006. On sample survey designs for consumer price indexes. *Survey methodology*, Vol. 32, No. 2 (December), p. 197-216.

EUROSTAT. 2001. *Compendium of HICP reference documents*. Bruxelles: European communities.

EUROSTAT. 2003. Outline of a Possible strategy for harmonizing HICP lower level aggregation. Paper presented at the meeting of the working party "Harmonization of consumer price indices", Luxembourg, 9-10 December.

Fenwick, D. 1998. The impact of sample design on the performance of the sample geometric mean and related issues. Paper presented at the fourth meeting of the International working group on price indices, Washington D.C., 22-24 April.

Fenwick, D. 1999. The impact of choice of base month on the relative performance of different formulae used for aggregation of consumer price index data at an elementary aggregate. Paper presented at the fifth meeting of the International working group on price indices, Reykjavik, 25-27 August.

Fenwick, D. 2001. Sampling in consumer price indices: what role for scanner data? Paper presented at the sixth meeting of the International working group on price indices, Canberra, 2-6 April.

Hansen, C.B. 2006. Price updating of weights in the CPI. Paper presented at the ninth meeting of the International working group on price indices, London, 14-16 May.

Koskimäki, T., M. Ylä-Jarkko. 2003. Segmented markets and CPI elementary classifications. Paper presented at the seventh meeting of the International working group on price indices, Paris, 27-29 May.

Kott, P.S. 1984. A superpopulation theory approach to the design of price index estimators with small sampling biases. *Journal of business & economic statistics*, Vol.2, No.1 (Jan).

ILO (ed.). 2004. *Consumer price index manual: theory and practice*. Geneva

ISTAT. 2009. Come si rilevano i prezzi al consumo. www.istat.it.

Leaver, S.G., J.E. Johnstone, K.P. Archer. 1991. Estimating unconditional variances for the U.S. consumer price index for 1978-1986. Proceedings of the Survey research methods section, American statistical association, p. 614-619.

Leaver, S.G., W.L. Weber, M.P. Cohen, K.P. Archer. 1987. Item-outlet sample redesign for the 1987 U.S. consumer price index revison. Invited paper 12.3 at the 46th session of the ISI-IASS meeting, Tokio 8-16 September. Booklet, p.278-286.

Leifer, H.-A. 2002. The 'elementary level' of a consumer price index and cost-of-living index: similarities and differences. Paper presented at the meeting of the Working party "Harmonization of consumer price indices", Luxembourg, 30 September-1 October.

Leifer, H.-A. 2008. Are the Carli, Dutot and Jevons indices time-path dependent? Paper presented at the fifth meeting of the HICP Task force sampling, Luxembourg, 19 February.

Norberg, A. 2004. comparison of variance estimators for the consumer price index. Paper presented at the eight meeting of the International working group on price indices, Helsinki, 23-25 April.

Norris, N. 1940. The standard errors of the geometric and harmonic means and their application to index numbers. *The annals of mathematical statistics*, Vol. 11,No. 4 (Dec.), p. 445-448.

Ribe, M. 2000. Price index, statistical inference and harmonisation. Proceedings of the fourth International conference on methodological issues in official statistics. 12-13 October.

Silver, M., S. Heravi. 2006. Why elementary price index number formulas differ: price dispersion and product heterogeneity. *IMF working paper*, WP/06/174.

Viglino, L. 2003. Harmonisation of HICP lower level aggregation. The use of the geometric mean for EEG and EA aggregation. Paper presented at the meeting of the Working party "Harmonization of consumer price indices", EUROSTAT, Luxembourg, 9-10 December.

Wilkerson, M. 1967. Sampling error in the consumer price index. *Journal of the american statistical association*, Vol. 62, No. 319 (Sep.), p. 899-914.