Enhancing the Quality of Price Index Estimates Combining Updated Weights, a More Representative Sample Design and a Different Aggregation Structure

Daniele Toninelli[1], Zdenek Patak[2], Martin Beaulieu[3]
[1]University of Bergamo, via dei Caniana, 2, 24127, Bergamo, Italy &
University of Ostrava, Sokolská třída 33, 70121, Ostrava, Czech Republic
[2] Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa ON K1A0T6, Canada
[3] Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa ON K1A0T6, Canada

Abstract
Consumer Price Index (CPI) and Service Producer Price Index (SPPI) calculation proceeds in stages. In the first stage, elementary aggregates of homogeneous sets of products are estimated. These elementary indices are then aggregated to obtain higher-level indices using, as weights, the elementary expenditure aggregates (CPI) or fiscal data from Business Register (SPPI). As regards SPPI, the robustness of the sampling design and the reliability of the index is often linked to an extended use of aging samples and basket weights in the aggregation process. By means of a simulation study based on data collected on business services and wholesale and retail trade, this paper aims at understanding how often weights should be updated to reduce the bias of the index. As regards CPI, in order to better represent the market shares of the various kind of stores selling same types of products, stratification by outlet type could be done. This raises questions as to what level would be more appropriate to define as the elementary aggregate. Our simulation study aims at assessing the impact on the index of the introduction of an additional intermediate level of weighted aggregation taking into account the stratification by type of outlet.

Key Words: Prices index, sample design, aging weights, index bias, Laspeyres index, Jevons index

1. Introduction

1.1 SPPI and CPI Overview
A price index is a measure of the proportionate change in a set of prices over time. In particular, in this paper two categories of price indexes are studied: *SPPI* (*Service Producer Price Index*) and *CPI* (*Consumer Price Index*).
The SPPI aims at measuring the price movements for the output of a business services (including wholesalers and retail traders). The target population of our SPPI study, in particular, is made of all businesses operating in Canada with at least one establishment within their structure coded to an in-scope *NAICS* (*North American Industry Classification System*) [1].

[1] For more information about the Statistics Canada's SPPI project, see Patak and Lothian (2007).

The CPI, on the other hand, is a measure of the monthly rate of change of the consumption prices of goods and services. To compile a CPI, a fixed basket of representative commodities is observed over time. The CPI basket is based on expenditures of a given target population (made of both families and individuals) living in private household in a certain reference period[2].
Both the SPPI and the CPI are based on the observation of prices of a basket of products/commodities over time. It is clear, at this point, that it is necessary to synthesize the variation observed for single prices by means of a single measure of the change (based on the group of representative products/goods sold by wholesalers or purchased by families and individuals). To summarize the evolution of prices over time, a wide group of options is available: the choice of the best price index formula to be used was object of intensive study and debate in the last years[3].

1.2 SPPI calculation

Even if we previously referred to a measure of prices' movements, the current methodology for the computation of the SPPI is based on the study of *profit margins* (*p.m.*s)[4]. Once observed two p.m.s referred to two consecutive months (month *t*-1 and *t*) and to a certain product (*k*) purchased and sold by a certain wholesaler's establishment (*i*), the simplest measure of the monthly variation is the ratio of the two observed profit margins:

$$\frac{p.m._{hik}^{t}}{p.m._{hik}^{t-1}},$$

where *h* is the stratum of establishment *i*. Starting from these ratios (*p.m.r.*s), a first level of aggregation is needed: for each establishment a predetermined number of products is observed (from 1 to 3 representative products), so we firstly need to synthesize the p.m.r.s at the establishment level. This is done computing a geometric mean of the p.m.r.s observed for each establishment. Doing so, we obtain the *elemental index*, computed with the Jevons formula[5]:

$$I_{hi}^{t,t-1} = \left[\prod_{k=1}^{m_{hi}} \left(\frac{p.m._{hik}^{t}}{p.m._{hik}^{t-1}} \right) \right]^{1/m_{hi}},$$

where m_{hi} is the number of products observed for the establishment *i* belonging to the stratum *h*. After the elementary level of aggregation, we need to further aggregate the

[2] For further details about the CPI program, see Statistics Canada (1995) and Beaulieu (2012).
[3] For best practices and further details, see the two ILO manuals (2004a, 2004b).
[4] A profit margin is computed as the difference between the selling price and the purchase price; see also, for more details, Patak and Lothian (2007). For more information and practical suggestions about the SPPI see, among others, the Producer Price Index Manual, Theory and Practice (ILO, 2004a).
[5] An intensive debate was made about the best index formula to be used at the elementary level for the aggregation: some interesting finding about this choice can be found in ILO (2004a); in more recent studies (Elliot *et al.*, 2012; Melser, 2005; Mehrhoff, 2010 and Dalèn, 1992 and 1999), through the stochastic approach, the choice of the best elementary aggregate formula is deeply discussed, linking it to the distribution of observed price/price relatives. Moreover, the paper of Balk (2005) deals with an alternative sampling approach, also oriented towards the identification of the best elementary index formula. At this regard it is also interesting the study developed by Hansen et al. (2003 and 2011), that follows the sampling approach and applies the Model Confidence Sets to different target indices to find out the best elementary aggregate level formula related to each target index.

obtained measures of change, computing an *aggregate index*, at the stratum (*h*) level. For this purpose, we use a Laspeyres index:

$$\hat{I}_h^{t,t-1} = \frac{\sum_{i=1}^{n_h}\left[z_{hi}^{t-1} \cdot w_{hi}^{t-1} \cdot I_{hi}^{(t,t-1)} \right]}{\sum_{i=1}^{n_h}\left(z_{hi}^{t-1} \cdot w_{hi}^{t-1} \right)}. \qquad (1)$$

This is a weighted average of the elementary indexes that uses two kinds of weights: the *sample weights* $\left(w_{hi}^t\right)$[6] and the *economic weights* $\left(z_{hi}^t\right)$[7].

1.3 CPI calculation

For the CPI a similar aggregation process is applied[8]. At the first step the price relative (*pr*) referred to a certain product *k* belonging to a stratum *h* is computed:

$$pr_{hk}^{t,t-1} = \frac{p_{hk}^t}{p_{hk}^{t-1}};$$

this price relative simply measures the relative change of the price of a certain good (*p*) between the month *t*-1 and the month *t*. At a later stage, we need to compute an elementary aggregate measure of the change: by means of the Jevons index, a geometric mean of price relatives referred to each stratum *h* is computed[9]:

$$I_h^{t,t-1} = \left[\prod_{k=1}^{n_h}\left(pr_{hk}^{t,t-1} \right) \right]^{1/n_h},$$

where n_k is the number of prices observed in the stratum *h.* The following step is the computation of the index at the top level of aggregation. For this purpose, we use a Laspeyres index formula[10]. This weighted average uses, as weights, the expenditures

---

[6] At a certain time *t*, the sampling weight $\left(w_{hi}^t\right)$ of a given establishment *i*, belonging to a stratum *h*, is computed, according to a certain sampling plan, as the inverse of the probability of being selected $\left(\pi_{hi}^t\right)$ of that establishment. For more information, see Toninelli (2010).

[7] The economic weight $\left(z_{hi}^t\right)$ for a given establishment *i* belonging to a stratum *h*, and referred to a certain time *t*, is computed as follows:

$$z_{hi}^t = \frac{\mathrm{Rev}_{hi}^t}{\sum_{i=1}^{N_h}\mathrm{Rev}_{hi}^t},$$

where $\left(\mathrm{Rev}_{hi}^t\right)$ is the annual revenue (source: BR) of a given establishment *i* (*i* = 1, 2, …, N_h) of a stratum *h* (N_h is ne number of establishments belonging to the stratum *h*).

[8] There are a lot of works dealing with the choice of the best formula to compute a CPI: see, among others, Armknecht and Silver (2012), Bishop (2013), Dalén (1998), Boskin et al. (1996). Some useful information and best practice can also be found in the CPI Manual (ILO, 2004b). Other interesting studies regard the relation between the two levels of aggregation: see, for example, Mehrhoff (2007, 2010).

[9] The stratum *h*, for CPI, is defined crossing the variables "elementary aggregate" and "geographic stratum".

[10] In our study about SPPI, we actually apply the Lowe index, that is a specific version of the Laspeyres formula in which weights come from a time *t* = *b* rather than from the base time *t* = 0 (*b* is the reference time for weights, that is the time of the last available update).

$\left(w_h^b\right)$ observed at a previous observed time, *b*, for stratum *h* (expenditure weights are updated every two years):

$$\hat{I}^{t,t-1} = \frac{\sum_{h=1}^{N_h}\left(w_h^b \cdot I_h^{t,t-1}\right)}{\sum_{h=1}^{N_h} w_h^b}. \qquad (2)$$

In this formula N_h is the number of strata.
Observing the formulas used at the top level of aggregation, for both SPPI (formula (1)) and CPI (formula (2)), we notice that the indexes are based, at the top level, on the computation of weighted averages. And both procedures are based on the use of specific weights: sampling weights (w_{hi}) and economic weights (z_{hi}) for SPPI and weights given by the expenditures shares (w_h) for CPI.

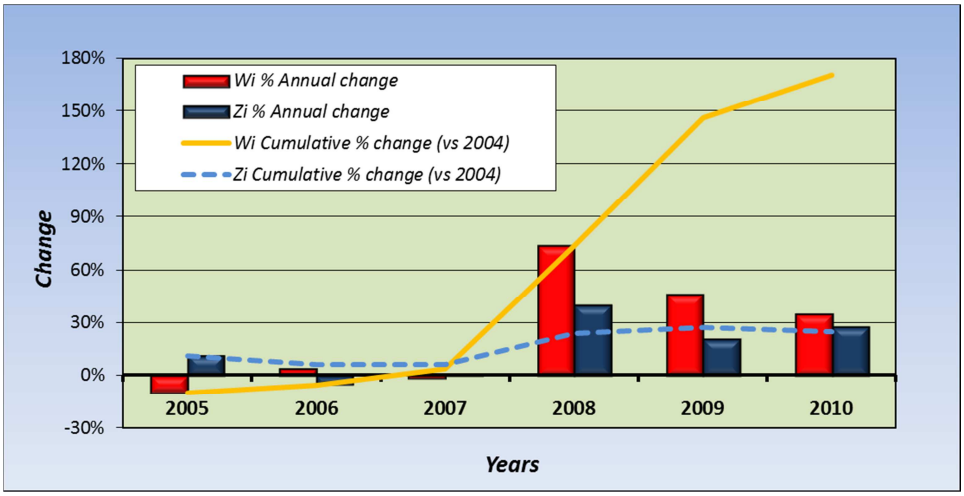
1.4 Objectives of the study

Our research, generally speaking, focuses on the study of issues related to the weighs used at the top level of aggregation and on their impact on price index estimates[11].
More in particular, for SPPI the problem of weights (that are basically computed on data coming from BR) is that they cannot be updated frequently: usually new BR data become available only with a certain delay. So our objective, for SPPI, is to understand, first, what is the impact of using aging weights on the reliability of the estimates, and, secondly, the combined effect with the sampling plan to detect which one can help in improving the quality of the estimates.
For the CPI study, our main objective is to understand if a different (and more detailed) weighted aggregation structure of the index can improve the quality of the final index. In particular, we want to understand if by adding an additional weighted aggregation level considering the different type of outlets the quality of the index could be improved, and the estimate can better represent the real purchases of Canadians. In regard to this, we consider two kind of outlets: the *Specialized stores* and the Non-specialized, that will be referenced, in the following, as *General stores*. This second group includes, among others, the NAICS 452110, 445110 and 444110.

2. Background and Methodology

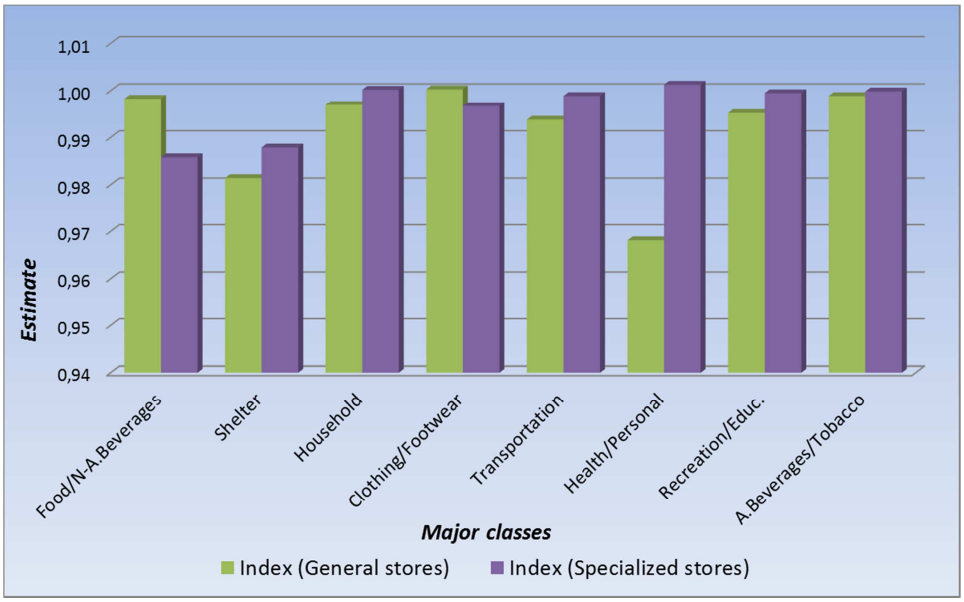
The objectives of our study were suggested by some evidences highlighted by a preliminary analysis.
More in particular, for SPPI we observed a noticeable variation of weights over time (Graph 1). This graph shows the annual (bars) and the cumulative (broken lines) change over time of sampling (w_i) and economic weights (z_i), between 2004 and 2010. During the 7 observed years, the average annual change for sampling weights is 14.2%; for economic weights it is 28.3%. Moreover, after 6 years, the cumulative change of sampling weight, in comparison to the first year (2004) is 170.2%, and the same change for economic weights is 24.9%.

[11] Interesting studies about the issue of weights, mostly in CPI, can be found in Greenlees and Williams (2009) and Hansen (2006). Mehrhoff also studied the properties of the two-staged price indexes (Mehrhoff, 2007) and the correspondence between elementary and aggregate index formulas (Mehrhoff, 2010). Further information can be found in the Producer Price Index Manual, Theory and Practice (ILO, 2004a) and in the Consumer Price Index Manual (ILO, 2004b).



Graph 1: Annual and cumulative percentage change of sampling (w_i) and economic weights (z_i) over time (2004 to 2010).

This analysis has also to deal with a change of the classification criteria of the BR frame, in 2008 (see big change in that year, underlined by the graph). Thus, the preliminary findings suggested our first research question for SPPI: can the use of aging weights lead to a low-quality general price index, significantly affecting the reliability of the estimates? And, moreover, how frequently weights should be updated to obtain a high-quality index and a more precise view of the general movements of prices?
The other evidence that encourages our study about the structure of the weight for the computation of CPI originated from the data represented in Graph 2. In this graph the CPI is computed for the General stores and for the Specialized stores categories: the final estimates for the two groups of outlets can be compared by major classes.



Graph 2: Comparison of the CPI estimates for General and Specialized stores strata (by major classes)

Graph 2 shows that there can be big differences in the estimate between the two type of stores. The biggest difference is observed for the Health/Personal care major class: the Specialized stores index shows a bigger price variation (+3.42%) than the General stores index. The minimum difference is detected for the Alcoholic beverages/Tobacco major class, where the Specialized stores index shows a slightly higher variation of prices (+0.1%). In other major classes (Food/Non-alcoholic beverages and Clothing/Footwear) the variation underlined by the General stores estimates is higher than the change underlined by the Specialized stores index.
Taking this all into consideration, the second research question (referred to the CPI weights structure) arises: can the introduction of an additional level of aggregation (taking into consideration weights specifically representing the General and Specialized stores categories) enhance the quality and/or the representativeness of the final estimates?

2.1 Data

The SPPI simulated population is based on data collected through the Statistics Canada's Wholesale Services Producer Price Index (WSPI), a program aimed at producing price deflators for the services sector. The target population for this survey is made of wholesalers' establishments operating in Canada; the data source for this population (and for the relative revenues) is the Business Register (BR). The population is stratified by NAICS (we used the 4-digit NAICS version) and by revenue (deciles of the distribution). The target population dataset includes the years from 2004 to 2010. The sample selection, for the WSPI survey, is based on two stages: at the first stage, using a Sequential Poisson Sampling (Patak and Lothian, 2007), a group of representative establishment is selected; at the second stage, from one to three products representative of the production of each establishment are selected (judgemental sampling). The original prices of these products are monthly collected.
The CPI study simulated population is also based on survey data collected monthly from January 2012 to January 2013. The dataset includes prices collected from a sample of products and services, representative of what Canadians consume. The sample design can be seen as a 3-stage sampling plan with non-probabilistic selection of representative items (Beaulieu, 2012). At the first stage a certain number of representative cities for each geographical stratum is selected (geographical component); at the second stage, for each one of the selected cities, a certain number of outlets where the targeted representative products and services can be priced are selected (outlet component); the third and last stage regards the selection of products, which can actually be found in the selected outlets, whose prices will be collected (product component)[12]. The primary source for weights data (expenditure shares) is Statistics Canada's Survey of Household Spending (SHS). The classification by type of outlets, on the other hand, is based on the North American Industrial Classification System (NAICS), available on the Statistics Canada's Business Register (BR) and their sales (or market shares) are based on Statistics Canada's Quarterly Retail Commodity Survey (QRCS).

2.2 Methodology

The methodology of this study, focused on both SPPI and CPI, starts from a common base: the generation of a simulated population of data. Another common aspect is the computation of different versions of price indexes that, later, are compared to each other. The SPPI approach is described, more into details, in par. 2.2.1, whereas the methodology of the CPI study is introduced in par. 2.2.2.

[12] For further details about the sampling plan, see Beaulieu (2012).

2.2.1 The SPPI simulation study

As previously said, the study of SPPI started from the generation of a simulated population. Based on micro-cells defined crossing two variables (4 digit NAICS, deciles of establishments' revenue), the best fitting distribution for the variable profit margins' relatives was detected and its parameters were computed. According to the results of this preliminary analysis, a simulated population of relatives was generated and, afterwards, linked to the frame of establishments[13]. Once linked the target population frame with the simulated profit margins relatives, following a bootstrap-like methodology (number of iterations: 3,000), we draw samples according to some of the main sampling schemes; in this paper only the results obtained with the following schemes are shown: Simple Random Sampling Without Replacement (SRS), Sequential Poisson Sampling (SPS), and Probability Proportional to Size (PPS) sampling. After the selection of each sample, two kind of indexes were computed on the same price relatives data: the first one, that we call *Current Index*, is the estimate obtained using current period's weights; the second one, defined *Simulated Index*, is computed using base period weights. This means, for example, that to compute the Current price index for a certain year x (x = 2004, …, 2010), the weights coming from the same year x are used. Whereas, to compute the Simulated index for the same year x, we used weights coming from a different year y (x, y = 2004, …, 2010; $x \neq y$). The evaluation criterion is based on the comparison of the Current index with the Simulated indexes estimates. Some of the results are also compared to the *Actual index*: this is the index computed, without sampling, on the whole target population, using the Lowe formula (that is a Laspeyres formula with weights referred to a time b). In our simulation study both the bias and the standard deviation of the final estimates are evaluated, considering the combination of the following two effects: a) the use of aging rather than updated weighs and b) the sampling plan effect combined with the use of updated weights (the main results are shown in par. 3.1).

2.2.2 The CPI simulation study

The study of the CPI is also based on a simulated population. The population of price relatives was generated following a methodology similar to the one introduced in par. 2.2.1. Nevertheless, there are two fundamental differences in the development of the simulation study. First, the population of price relatives was generated starting from the study of the distribution by micro-cells, similarly to what done with the SPPI, but the micro-cells were defined by elementary aggregate, geographical strata and reference month. The second difference is that the CPI is computed on a large part of the target population; that is, the population object of the study is the sample of representative goods (basket) whose prices are collected by Statistics Canada. Only elementary aggregates for which prices are obtained via an alternate data source (e.g. administrative data) are excluded from the scope of this study. The CPI simulated study strategy is mainly focused on the evaluation of the effect obtained introducing an additional intermediate level of weighted aggregation (based on the type of store considered). For this reason the analysis is made comparing different alternative of estimates computed on the same dataset. The different version of compiled indexes are the following:

- *Classic Index* (*CI*): it is computed using the "classic" current methodology (Jevons index at the elementary aggregate level, Laspeyres weighted formula to aggregate indexes all the way up to the all-item index).

[13] The frame includes all active establishments listed in the BR dataset of Canadian wholesalers (nevertheless the smallest units were deleted; for further details, see Patak and Toninelli, 2009, Toninelli, 2010, and Toninelli and Patak, 2010).

- *General sub-Index* (*GsI*): it is the sub-index computed (with the classic methodology) on the General store group of outlets[14];
- *Specialized sub-Index* (*SsI*): sub-index computed with the classic methodology on the group of Specialized stores;
- *Final Index* (*FI*): this index is computed introducing the intermediate weighted level of aggregation; we obtain four versions of this index:
    - *Final Arithmetic* (*FA*): this index is computed as an arithmetic weighted average of GsI and SsI, using, as weights, the type of store market shares;
    - *Final Geometric* (*FG*): it is computed as a geometric weighted average of GsI and SsI (weights: market shares);
    - *Final Arithmetic Low* (*FAL*): a first weighted aggregation is made at the low level (elementary aggregate level) by type of stores, using an arithmetic weighted formula (weights: market shares by type of store);
    - *Final Geometric Low* (*FGL*): the computation of the index introduces a geometric weighted average (weights: market shares by type of store) at the low level of aggregation (elementary aggregates);

The final estimates obtained with the above mentioned strategies are compared and discussed in par. 3.2.

3. Results

In this section the main findings obtained studying the SPPI (par. 3.1) and the CPI (par. 3.2) are discussed.

3.1 Service Producer Price indexes

Our work was first focused on the study of the impact of aging weights in the SPPI framework. In particular, we evaluated the combined effect of aging weights and sampling plans on the final estimates, comparing the Simulated and the Current indexes with the Actual index estimates. In this paper the results obtained applying the SRS, the SPS and the PPS sampling are compared.

If samples are selected through a SRS without replacement, the bias of the estimates (Simulated indexes) is highly unstable if we use different weights, but it is usually smaller than the bias obtained with a probability proportional to size sampling plan (such as SPS and PPS). On the other hand, the standard deviation is quite stable (independently from the used weights), but it could be up to 50 times the Current index's standard deviation. If we select samples using a probability proportional to size sampling plan we usually obtain an unstable bias (highly depending from the weights used to compute the Simulated index, and strongly linked to the current year too); moreover the bias is bigger than the one obtained using SRS. However, this is expected, seen that probability proportional to size schemes select with a higher probability the bigger units (and these are more likely to show bigger variations of profit margins). Generally speaking, there are no substantial differences between results obtained with the SPS and with the PPS sampling. Nevertheless, if compared to the SRS, these sampling schemes allow to obtain a smaller standard deviation (at an average, it can be up to 3.08 times the standard deviation of the Current index). This is a positive aspect, seeing that a smaller standard deviation means that the final estimates have higher precision.

[14] The General stores category is better defined in Section 1.4.

In our study we also study the combined effect of sampling plan and aging weight on both the bias and the standard deviation of the final estimates. In this regard, Table 1 shows the percentage variation of bias and standard deviation, by sampling plan, that is obtained using updated rather than not updated weights; the results are given by the comparison of the Current and of the Simulated indexes' statistics with the statistics of the estimates obtained from the whole population (Actual index).

Table 1: Percentage variation of bias and std. dev. of the estimates using updated weights (bootstrap-alike sample selection by sampling plan; 3,000 iterations)

	Sampling plan		
	SRS	*PPS*	*SPS*
Bias	-31.5	+0.4	+5.9
Std. dev.	-0.4	-16.6	-18.1

It was already said that the SRS produces estimates with a smaller bias; moreover this can be strongly reduced using updated weights (-31.5%); nevertheless the variance of the index is highly inflated (and this means that the estimates are not precise), and using updated weights the standard deviation is not noticeably reduced (only -0.4%). Probability proportional to size schemes, on the other hand, bring to estimates with a bigger bias[15], that is not significantly increased (+0.4%) or only slightly increased (+5.9%) using, respectively, PPS and SPS. If there is no gain in term of bias, using updated weights the standard deviation of the index could be strongly reduced: -16.6% with PPS and -18.1% with SPS.

3.2 Consumer Price Index

This section will discuss some results about the introduction of a new intermediate weighted level of aggregation in the index compilation scheme first (par. 3.2.1); secondly, the main findings of a deeper study of the evolution of the estimates, also by major classes, are shown (par.3.2.2).

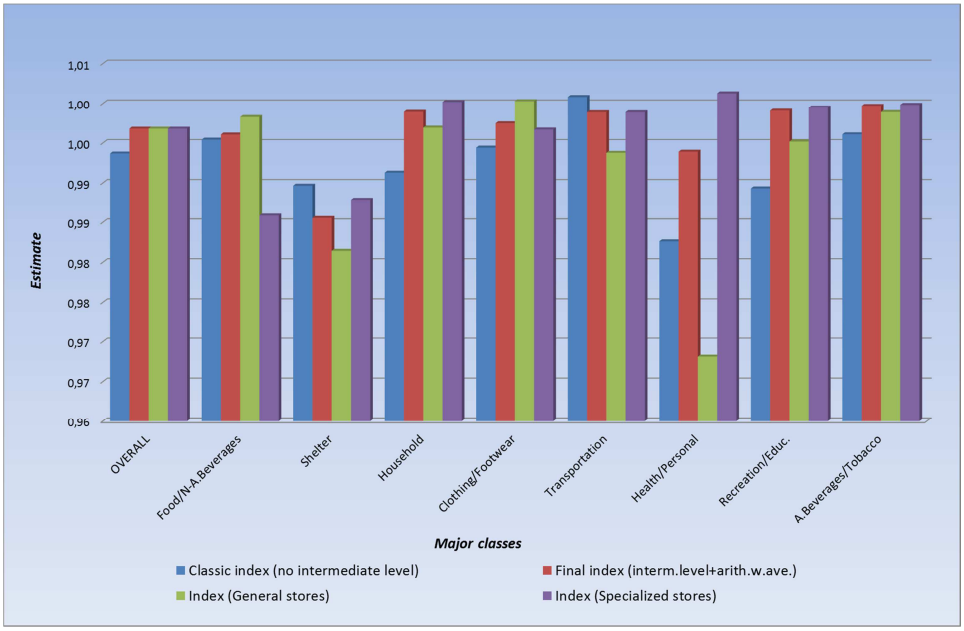
3.2.1 Introduction of an intermediate level of aggregation

Graph 3 shows the comparison of four indexes: the CI (obtained with the "classical" methodology of aggregation, i.e. Jevons index+Laspeyres index), the FI (in the graph the Final Index computed with the weighted arithmetic average is shown), and the two sub-indexes regarding, respectively, the general and the specialized type of stores (GsI and SsI, respectively). The comparison is made at a general level (that is considering the whole frame population; see the first category on the left of the Graph 3) and by major classes (from the second to the last classes on the *x*-axis of the same graph).

Overall (first class of the Graph 3), we notice that the Final index (FA) and the sub-indexes show very similar results; the Final index is also extremely coherent with the two boundaries represented by the General and Specialized stores indexes. On the opposite side, the Classic index tends to measure a noticeably lower level of prices' movements. The behavior of the two sub-indexes seem identical, at the overall level, but taking into consideration the different major classes, we notice big differences in the sub-indexes' estimates. For all the categories (but for Clothing/Footwear) the GsI shows lower estimates, whereas the SsI shows higher values. The FA is always included in between

[15] This is the effect of the bigger weights given to the biggest units of the sample: they usually show a higher variation of prices over time (Toninelli, 2010).

the two sub-indexes boundaries, as expected. But from this graph also emerge the fact that the CI is not coherent with the other indexes, and mostly to the sub-indexes. In several cases (6 out of eight and at the overall level) the CI is out of the two sub-indexes boundaries: it is included between the two sub-indexes boundaries only for Food/Non-alcoholic beverages and for the Health/Personal care major classes. In 2 cases the CI give bigger estimates than all other indexes (for Shelter and Transportation), in other 4 cases it provides smaller values (for Household, Clothing/Footwear, Recreation/Education, and Alcoholic beverages/Tobacco).

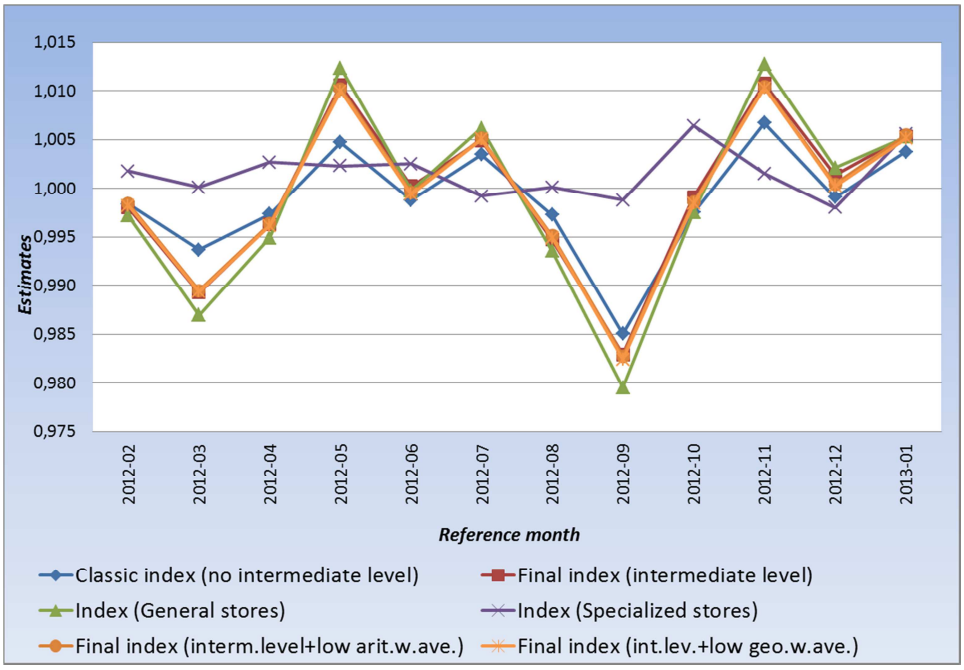


Graph 3: Classic index, Final index and General and Specialized stores sub-indexes comparison (overall and by major classes)

Graph 3 only shows the FA index: the FG index is not shown because the differences between the estimates obtained with the geometric and the arithmetic formulas are not noticeable. Also the FAL and the FGL indexes are not shown in the graph, but they confirm each other (the index obtained using the geometric formula is always slightly smaller than the index obtained with the arithmetic formula); moreover, they are substantially coherent with the results obtained with the FA and the FG aggregation structures, at the overall level: they only provide slightly smaller estimates of the price movements (except for the Shelter major class, where they give noticeable higher estimates than the levels provided by the FA and the FG indexes). But what is more important is that the FAL and the FGL estimates almost always (but for Shelter and, regarding the FGL, also for Clothing/Footwear) fall within the sub-indexes boundaries.

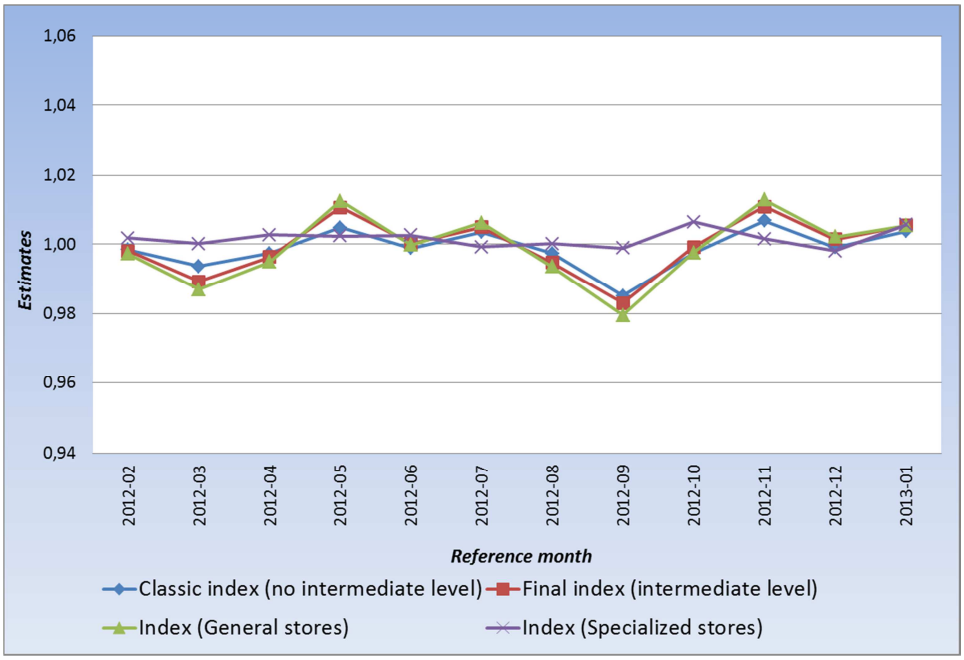
3.2.2 Estimates over time

In Graph 4 the evolution of the indexes over time is shown: the CI (light blue) is directly compared to FA (red), to the sub-indexes (GsI, in green, and SsI, in purple), and to the FAL and the FGL indexes (in light brown and orange, respectively).

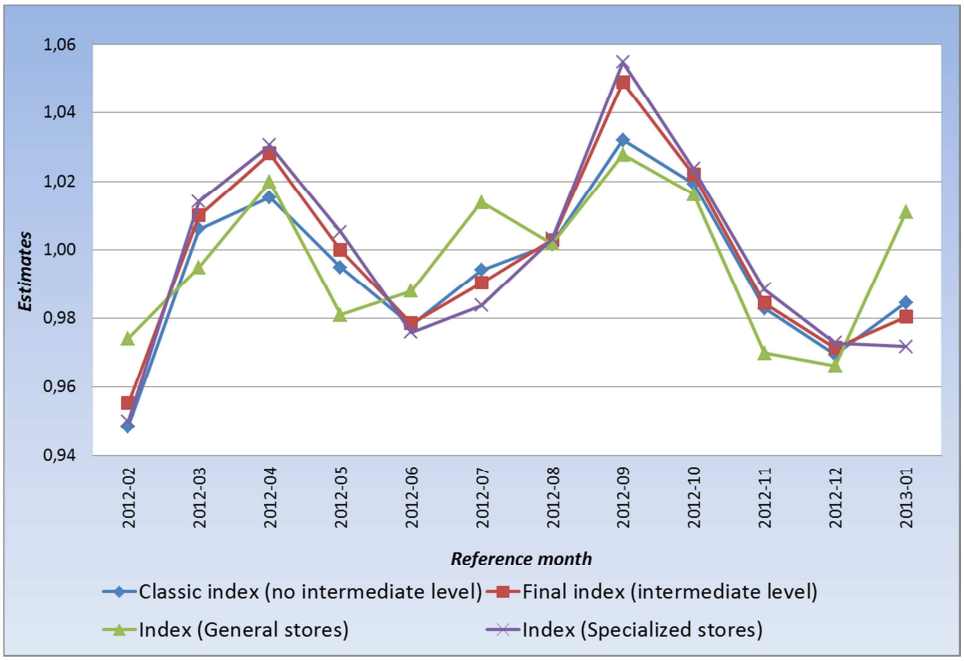


Graph 4: Comparison between indexes over time (Feb. 2012 to Jan. 2013, all classes)

Overall, we notice a high coherency in the estimates given by the Final Indexes (at both the top and low level of aggregation, with no differences given, at the low level, by the kind of aggregation formula). All Final indexes, moreover, are mostly influenced by the behaviour of the General stores index. The CI is, on the other hand, the aggregated index closer to the (more stable) behaviour of the Specialized sub-index. This graph also shows that it is possible to obtain different results using the two different aggregation structures.
Taking a more detailed look into two of the more interesting major classes (Food/Non-alcoholic beverages and Clothing/Footwear) we can better understand what should be the behaviour of our "desired" estimates. If we take into consideration the Food/Non-alcoholic beverages major class, we know that the percentage of general stores, in terms of number of observed prices, is 97.03% (the same percentage in terms of market shares is 82.30%). Thus, taking into account these weights, we would expect a representative index proportionally more influenced by the behaviour of the General sub-index than by the Specialized index. The weighted Final indexes, in fact, follow more the General than the Specialized one (see Graph 5), whereas the CI is more a compromise with the more stable SsI.
The Clothing/Footwear major class (Graph 6) is more well-balanced in terms of percentage of observed prices (49.67% for General stores and 50.33% for Specialized stores). Nevertheless, the percentages, in terms of market shares, is noticeably higher for Specialized stores (78.28% vs 21.72% of General store). Thus, for this major class we can assume that a price index that follows more closely the Specialized stores' index behaviour would be more representative; and this happens for the Final indexes, whose strictly follow the SsI, whereas the CI is more influenced by the GsI, and, sometimes (February and April 2012), it is also out of the sub-indexes' boundaries.



Graph 5: Comparison between indexes over time (Feb. 2012 to Jan. 2013); major class: Food/Non-alcoholic beverages



Graph 6: Comparison between indexes over time (Feb. 2012 to Jan. 2013); major class: Clothing/Footwear.

4. Final remarks and recommendations

4.1 About SPPI

The study of the use of aging weights shows that we can obtain an over rather than an underestimation of the AI and a level of variance very different year by year. This general

conclusion is confirmed by similar results obtained studying clusters of NAICS. Our results suggest that we should update the weights, in compiling the SPPI, as frequently as possible. Ideally weights should be updated annually, but this is almost always impossible, seeing the lag necessary to obtain new data (and, therefore, new updated weights) from the BR. If we use weights older than two years, at an average level, we obtain a bigger variance and a higher standard deviation of the estimates. But this does not always happen: weights do not follow a linear trend, so sometimes, in compiling SPPI, we can be lucky and we can use weights coming from a very far year that has a structure more similar to the one of the current year than that one of a closer year. The most important thing that has to be highlighted, following our results, is that the variance of the index estimates is usually higher, if not-updated weights are used. Among all analyzed sampling plans, Probability Proportional to Size sampling schemes seem to be able to provide more reliable estimates (noticeably reducing their standard deviation), mostly if updated weights are used.

4.2 About CPI

The "weights" topic for CPI is also of fundamental importance. In this paper we observed that the sub-indexes obtained considering the type of outlets (General/Specialized) are clearly different. Thus, the introduction of an additional weighted level of aggregation between the low level (Jevons index) and the high level (Laspeyres index), according to some preliminary findings of our research, push us to suspect that we could obtain more representative estimates, taking into account the different contribution (in terms of market shares) of the two kind of outlets.

In this paper we compared different methodologies to compile a new version of the aggregated indexes (Final indexes). We mainly compared the arithmetic and geometric formula to aggregate the sub-indexes at the aggregate level, but no significant differences in the obtained results were highlighted. Similar results were also observed comparing the FAL with the FGL scheme of aggregation, where, respectively, the arithmetic and the geometric weighted average are computed at the elementary level of aggregation. Our study also shows a high coherency of the four versions of the FIs (they mostly follow a common pattern of evolution) and a coherency with the two boundaries represented by the two sub-indexes (referred to General and Specialized stores). Most of all, we found that the CI (compiled with the classical methodology) is not so coherent with the other versions of the indexes (and most of the times it is also not consistent with the two sub-indexes boundaries). The general conclusions seem to be confirmed by a study developed at the major classes level (considering, for example, two of the main ones: the Clothing and the Food/Non-alcoholic beverages).

Generally speaking about the obtained results, we also have to underline the current limits of our work. We cannot conclude, at this point, if the Final index aggregation structure (introducing the intermediate weighted level of aggregation) is better (or brings to more representative results) than the "classical" methodology. We only discovered that there are noticeable differences in the results, using the two methods; we also obtained some indication (coherency and evolution over time) that seem to confirm that the Classic way of compiling a CPI could have some weaknesses in comparison to the new proposed methodology, especially when the sample distribution between General and Specialized stores does not reflect their respective market shares. Thus, further research is extremely needed to get a confirmation of our initial findings.

4.3 Further research

Further research is still needed about weights used for the SPPI and about the weighted aggregation procedure applied to compile the CPI.
For SPPI, in particular, we need to understand if there is some way to identify a more clear trend in the weights' evolution over time, studying a wider range of data (that is, considering more years). Further research is also needed, in this perspective, as soon as new BR data about the target population will be available, eventually extending the study at the economic sector level (a more detailed view could bring to a clearer picture of the evolution of both prices and weights over time).
For CPI a further plan of study is already under development, starting from these first results. Computing chained versions of the indexes studied in this paper, we want to estimate the cumulative effect (over years) of the introduction of the intermediate level of aggregation on the final estimates. Moreover, in order to face some of the limitations of our study, we planned to compare the Classic and the Final Indexes's estimates to a "superlative index" (e.g., using the Fisher or the Törnqvist formulas). The results obtained with the classical method and with the new proposed methodology could be then compared to a benchmark, considering a longer interval of years and, eventually, taking also into account the "change of the basket" effect and its impact on the reliability of the final estimates.

Acknowledgements

This research project has been developed with the support of the FYRE (Forecasting Young REsearcher) project funds (Fondazione Cariplo/University of Bergamo, 2011-13). The research was also funded by the project "Opportunity for young researchers", reg. no. CZ.1.07/2.3.00/30.0016, supported by Operational Programme Education for Competitiveness and co-financed by the European Social Fund and the state budget of the Czech Republic. The paper has been also funded by the 60% university funds (University of Bergamo).

The authors have to thank Catalin Dochitoiu (Statistics Canada) for his essential contribution and for his precious suggestions in developing this project, and also for his always being a very kind and helpful colleague. We also have to thank our colleagues of the BSMD (Statistics Canada) for their useful suggestions and comments.

Disclaimer

The contents of this article reflect the views of the authors and not necessarily the official views or opinions of Statistics Canada.

References

Armknecht P., Silver M. (2012). Post-Laspeyres: The Case for a New Formula for Compiling Consumer Price Indices. IMF Report WP/12/105.

Balk, B.M. (2005). Price Indexes for Elementary Aggregates: The Sampling Approach. Journal of Official Statistics, 21, pp. 675-699.

Beaulieu, M. (2012). Methodological Challenges in the Development of a New Survey Design for the Canadian Consumer Price Index. In ICES IV - International Conf. on Establishment Surveys Proceedings, Montréal.

Bishop, G. (2013). Geometric Indexes and Substitution Bias in the CPI. Ottawa Group Meeting 2013 proceedings. Australian Bureau of Statistics.

Boskin, M.J., Dulberger, E., Gordon, R., Griliches, Z., and Jorgenson, D. (1996). Toward a More Accurate Measure of the Cost of Living. Final Report to the Senate Finance Committee from the Advisory Commission To Study The Consumer Price Index (December 4th).

Dalén, J. (1992). Computing Elementary Aggregates in the Swedish Consumer Price Index. Journal of Official Statistics, 8 No. 2, pp. 129-147.

Dalén, J. (1998). Studies on the Comparability of Consumer Price Indices. International Statistical Review, 66, pp. 83-113.

Dalén, J. (1999). A Note on the Variance of the Sample Geometric Mean. Department of Statistics, Stockholm University, Research Report 1.

Elliott, D., O'Neill, R., Ralph, J., and Sanderson, R. (2012). Stochastic and Sampling Approaches to the Choice of Elementary Aggregate Formula. Office for National Statistics, Discussion Paper.

Greenlees, J.S., and Williams, E. (2009). Reconsideration of Weighting and Updating Procedures in the US CPI. Working paper 431, June 5, 2009. U. S. Bureau of Labor Statistics

Hansen, P., Lunde, A., and Nason, J. (2003). Choosing the best volatility models: the model confidence set approach. Oxford Bulletin of Economics and Statistics, 65, pp. 839-861.

Hansen, C.B. (2006). Price Updating of Weights in the CPI. 9th Meeting of the International Working Group on Price Indices, London, UK.

Hansen, P., Lunde, A., and Nason, J. (2011). The model confidence set. Econometrica, 79, pp. 453-497.

International Labour Office (ILO), IMF (2004a). Producer Price Index Manual, Theory and Practice. Washington, D.C.: International Monetary Fund.

International Labour Office (ILO), IMF, OECD, Eurostat, UNECE, and the World Bank (2004b). Consumer Price Index Manual: Theory and Practice. Geneva: International Labour Office.

Mehrhoff, J. (2007) Aggregate Properties of Two-Staged Price Indices.

Mehrhoff, J. (2010). Aggregate Indices and Their Corresponding Elementary Indices. Journal of Economics and Statistics Band 230 Heft 6, Lucius & Lucius: Stuttgart.

Melser, D., (2005). The Choice of Elementary Index Formula in the UK CPI/RPI. Internal ONS report.

Patak, Z. and Lothian, J. (2007). Enhancing the Quality of Price Indexes, a Sampling Perspective. Ottawa, ON: Statistics Canada.

Patak, Z., and Toninelli, D. (2009). Are prices surveys sample designs robust to aging weights? A simulation study. Proceedings of the XXV International Methodology Symposium 2009, Gatineau, QC. Ottawa, ON: Statistics Canada.

Statistics Canada (1995). The Consumer Price Index Reference Paper Update based on 1992 Expenditures. Catalogue 62-553: Statistics Canada.

Toninelli, D. (2010). Survey Techniques for Price Indexes: An application to prices data for the computation of price indexes. Saarbrücken, Germany: VDM Verlag.

Toninelli, D., and Patak, Z. (2010). The Effect of Weights and Sampling Plans on Price Index Estimates: a Simulation Study. In 2010 JSM Proceedings, Vancouver, BC. Alexandria, VA: American Statistical Association.