Proceedings of Statistics Canada Symposium 2009
Longitudinal Surveys: from Design to Analysis

Are Prices Surveys Sample Designs Robust to Aging Weights? A Simulation Study

Zdenek Patak, and Daniele Toninelli[1]

Abstract

The importance of the prices' study has increased quickly in the last years: several national statistical agencies started developing projects based on longitudinal studies to measure the movements of the prices of products and services. Many methodological issues are subject of intensive research to improve the quality of the whole indexes' production process. The target of this work is to contribute to improve the quality of the data collection process by studying the temporal evolution of the survey data. Starting from a simulated population, a number of commonly used sample selection methods are compared underlining their relative efficiency and evaluating how and how much the change in the size measure over time affects the estimates and the bias of the results.

Key Words: Prices Index, Sampling Weights, Simulation Study, Sampling Methods, Index Bias.

1. Introduction

With the release of the Boskin Report (Boskin *et al.*, 1996) on the state of U.S. price indexes, there has been an intense debate on ways to improve their quality. Major strides have been made in assessing the applicability of standard business survey methodologies in an area that has long been dominated by judgment and subject-matter expertise. In the U.S., in particular, the coverage, collection and sampling processes have experienced major innovations. With the growing importance of the service sector, Statistics Canada is developing a new set of Service Producer Price Indexes that may incorporate many of these innovations. Statistics Canada is also using this opportunity to investigate issues such as the impact of sampling design on the precision of an index, and the effect of prolonged use of aging weights on reliability. By means of a simulation study, this paper compares a number of probability sampling strategies, such as simple random sampling without replacement (SRSWOR), and several probability proportional to size (PPS) methods, commonly used in practice in the context of estimating a price index. The study was done to assess the robustness of various designs to aging samples and basket (a set of goods or services whose prices are being monitored over time) weights in terms of bias and mean squared error of the estimated index. The target population for the simulation study was generated based on parameters computed from the Canadian Wholesale Services Producer Price Index survey.

1.2 Price index computation

The first step in obtaining the final index is the computation of the *elemental index* at the establishment level (lowest level of aggregation). Let p_{hik}^{t} be the profit margin observed at time *t* for product *k* (k = 1, …, m_{hi}) of establishment *i* belonging to group *h*, where m_{hi} is the number of observed products for establishment *i* (m_{hi} = 1, 2, 3). The elemental index $\left(P_{hi}^{t/t-1}\right)$ between times *t* and *t*-1 is computed as the geometric mean of the profit margin ratios,

$$P_{hi}^{t/t-1}=\left[\prod_{k=1}^{m_{hi}}\left(\frac{p_{hik}^{t}}{p_{hik}^{t-1}}\right)\right]^{1/m_{hi}}. \qquad 1.2\text{-}1$$

This index is also called *Jevons index*[2]. The *aggregated index* at the stratum level of aggregation (stratum *h*) is a weighted average of elemental indexes. It is computed using the following Laspeyres formula[3],

[1]Zdenek Patak, Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa, Ontario, Canada, K1A 0T6 (Zdenek.Patak@statcan.gc.ca); Daniele Toninelli, University of Bergamo, via dei Caniana 2, Bergamo, Italy, 24127 (daniele.toninelli@unibg.it).
[2] See the *Producer Price Index Manual* (2004) and the *Consumer Price Index reference paper* (1995).
[3] For further information about the Laspeyres index see: the *Producer Price Index Manual* (2004), the *Consumer Price Index reference paper* (1995), and Laspeyres (1884). For more details about the use of the Laspeyres index in the SPPI project context, see also Patak and Lothian (2007).

$$\hat{I}_h^{t/t-1} = \frac{\sum_{i=1}^{n_h} \left[z_{hi}^{t-1} \cdot w_{hi}^{t-1} \cdot P_{hi}^{(t/t-1)} \right]}{\sum_{i=1}^{n_h} \left(z_{hi}^{t-1} \cdot w_{hi}^{t-1} \right)}, \qquad 1.2\text{-}2$$

where h is stratum, n_h is the number of units of stratum h, z_{hi}^{t-1} is *economic weight* of unit i (belonging to the stratum h) at time t-1, and w_{hi}^{t-1} is *sampling weight* of unit i at time t-1. More generally speaking, for a year t, the economic weight z_{hi}^{t} of an establishment i belonging to the group of units h is computed as the ratio of the revenue of the considered establishment $\left(R_{hi}^t\right)$ to the total revenue of all the establishments belonging to the same group $\left(R_h^t\right)$: $z_{hi}^t = R_{hi}^t / R_h^t$. The sampling weight w_{hi}^t of a unit i for the year t is computed as the inverse of the probability of selection, $w_{hi}^t = 1/\pi_{hi}^t$. The probability of selection $\left(\pi_{hi}^t\right)$ of unit i is the probability of being included in the selected sample; it varies according to the sample selection methodology. In the following paragraph (1.3), we will test if sampling and economic weights vary over time. We will also try to answer the question: how often should weights be updated to get a value of the estimated index as close as possible to the value of the actual index?

1.3 How weights change over time

Both sampling and economic weights appear in the aggregated index formula (1.2-2), so potentially they both have an influence on the estimates. In fact, if a price index has to be computed, the system of weights to use has to be decided and this can affect the quality of the final result. The best choice for computing a price index for a certain year would be to use that year's weights: this would provide what we call an *actual index*; however this is not always possible as the actual weights are usually available with a delay. This means that weights that are not current are often used for computing a price index. Furthermore, it is important to understand if and how the weights vary over time, and what is the impact of their change on the bias and on the variance of the index. This paragraph studies the first aspects, while the impact of changing weights on the index is discussed in par. 3.1.

An analysis of the sampling (w_i) and economic weights (z_i) for the period from 2004 to 2009 showed that in the first three years the sampling weights remained relatively unchanged; the percentage variations stayed within 10%. The year over year variation started changing appreciably after four years, +73.6% between 2007 and 2008 and +46.1% between 2008 and 2009. If the first variation could be explained by a change in how the BR was structured, the same reason does not explain the 2008-09 change. The average annual change in sampling weights was +15.2%. Furthermore, the changes in sampling weights in the short term did not seem to follow a predictable path, i.e., the trend was negative between 2004 and 2005 and between 2006 and 2007, positive in the other cases. If we consider cumulative variations, there was an increase of 145.9% in the five years beginning in 2005 in comparison with 2004; the average cumulative change over the same period was 29.2%. This broad change in the sampling weights can understandably have a strong effect on the quality of the estimates if we are using extremely outdated weights.
The economic weights showed a similar pattern over the same six years (2004-2009), but with smaller peaks of variation compared to the sampling weights; the maximum year-over-year change was 39.8% between 2007 and 2008 (the impact of switching to the new BR is underlined again). The average annual variation was bigger (20.1%) than the one observed for the sampling weights.

The data show that sampling weights vary more in the long run, while economic weights vary more on an annual basis. Having said that, even if the biggest change (2007-08) is in part explained by an external event (that does not explain completely the 2008-09 change), we can conclude that these variations can potentially have a considerable impact on the precision of the final estimates. They can also reinforce themselves, given that both sampling and economic weights appear in the index formula. The next paragraph will introduce the methodology we used to evaluate the impact of changing weights (that is of weights getting aged) on the estimates of the index in terms of both bias and variance.

2. Methodology

Our study is based on a simulated population of price relatives. They were generated using parameters computed using price relatives observed in the 2006 to 2008 waves of the Statistics Canada's wholesale survey. The simulated relatives were then associated to the real frame population of establishments used to select the wholesale survey samples. Structural and fiscal data for each unit were obtained from the BR. Once the simulated frame population was generated, we selected samples using the following sampling designs: *Simple Random Sampling* (SRS), *Probability Proportional to Size* sampling (PPS), and *Sequential Poisson Sampling* (SPS). The selection of samples was carried out with a bootstrap-like methodology, selecting a maximum of

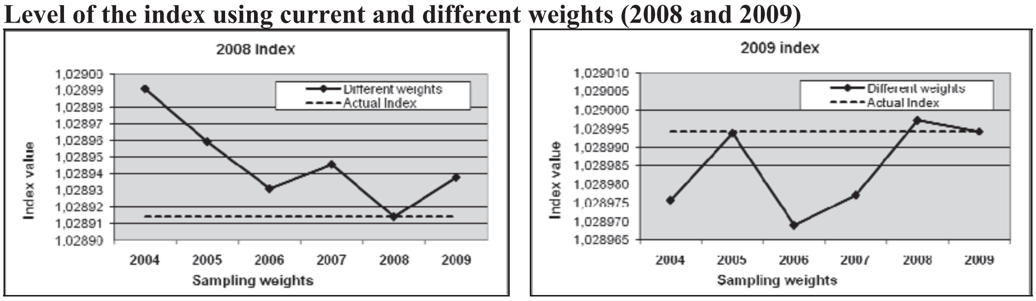
5,000 samples of 1,000 units each, in order to obtain convergence of the index' variance. For each of the selected samples we computed different versions of the index, varying the sampling weights (that is using sampling weights coming from all the BR' variables' considered years). The actual index was computed for each year of the simulated population (from 2004 to 2009) using the current period weights; this means that, for a year *x* (*x* = 2004, 2005, …, 2009), the system of weights of the same year (*x*) was used in computing the index. Furthermore, other *simulated indexes* were computed for each year, using weights coming from the other years: for example, to compute a simulated index of a year *x*, the weights of a year *y* (*y* = 2004, 2005, …, 2009, $y \neq x$) were used. For each version of the simulated index we also computed the variance and the bias obtained by comparing the simulated index itself with its actual value (that is the estimate obtained using current weights, i.e. weights belonging to the same year of the index). Thus, this bias stems from the use of not-updated weights. In this way, for each year's index we can evaluate both the bias of the estimates and the variance that we obtain using aging weights: this all gives us the evaluation of the estimates' precision. Furthermore, we can compare the results obtained for different sampling selection schemes, measuring their relative effect on the index bias as well as their robustness. The following part of the paper (par. 3.1) will focus, in particular, on the study of the effect of sampling weights (our experimental factor) on the estimates of the index. In the par. 3.2 we will discuss some results (in terms of both bias and variability) obtained with the implemented sampling selection methods. The comparative evaluation of such methods will follow (par. 3.3).

3. Some results

3.1 Level and variance of the index

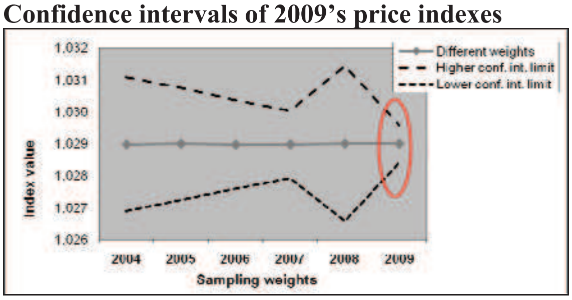
In this paragraph we evaluate the impact of sampling weights on the level of the price indexes. Using the whole frame simulated dataset, we computed, for each year (from 2004 to 2009), the actual index and the simulated indexes. In the graphs 3.1-1/2 the results about the 2008 index (3.1-1, on the left) and about the 2009 index (3.1-2, on the right) are shown. In both the graphs the broken line represents the actual index (that is the index computed using current sampling weights), while the black line shows the estimates of the simulated indexes (computed using not-current sampling weights).

Graph 3.1-1/2
Level of the index using current and different weights (2008 and 2009)



The most common situation (observed in 2005, 2006 and 2007) is similar to the one shown for the 2008 index (graph 3.1-1): the more we use weights that are far from the current year, the more the bias of the simulated index increases, in comparison with its actual value (2008 index computed with 2008's weights). However, sometimes (e.g. in 2004 and in 2009) we can also obtain a situation similar to the one underlined in the 2009's graph (3.1-2): even if we are using aged weights (i.e. weights of 2005) we can obtain estimates close to the value of the actual index. The graph 3.1-3 shows the confidence intervals built around the 2009 index values, obtained using different sampling weights. The depicted situation is common to the majority of the cases (2008 excluded): when we are using current weights, the computed index is more precise (that is, the confidence interval is thinner).

Graph 3.1-3
Confidence intervals of 2009's price indexes

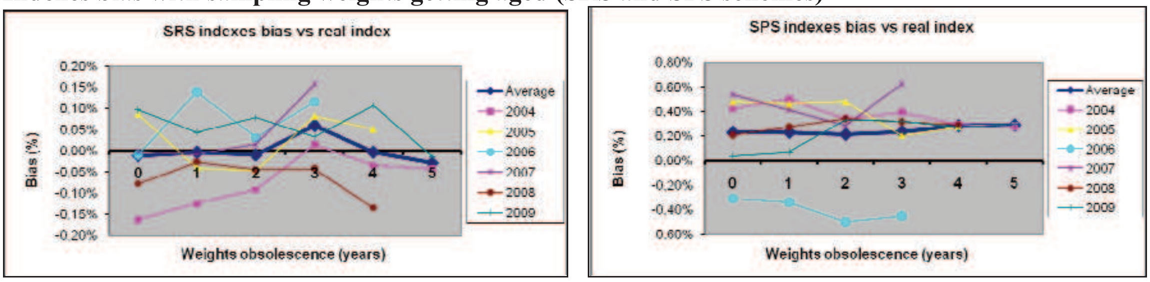


All this seen, it is possible to conclude that we do not always obtain extremely biased estimates when we do not use current weights, but usually these estimates are less precise than the ones obtained using updated weights. After analyzing what happens when the sample (and the sampling weights) become dated, in the next par. 3.2 we study the effect of the sampling methodology on the precision of the index.

3.2 Bias and standard deviation of the estimates

The results shown in this section are referred to a maximum number of 500 bootstrap iterations in selecting samples of 1,000 units from the simulated frame population. For the selection of the units the following methods were used: SRS, PPS and SPS. Seen that the results obtained with the PPS are similar to the one obtained with the SPS, only these last will be discussed in this paper. In the graphs 3.2-1/2 the bias obtained with the SRS (graph 3.2-1, on the left) and SPS (graph 3.2-2, on the right) is shown. In particular, we show the behaviour of the index while sampling weights are getting dated: on the *x* axis the distance (from 0 to 5 years) of the used weights from the year object of analysis is represented (so "0" means that we are using current weights).

Graph 3.2-1/2
Indexes bias with sampling weights getting aged (SRS and SPS schemes)



In the graph 3.2-1 the thick line represents the average path of all indexes while sampling weights are getting dated: the bias (and its variance) seems to increase more as we use weights that are far from the current year. However, we notice that the average situation is only the compensation of completely different paths shown from the thinner lines (referred to each single index). In fact, if we take a look at the details, we notice that for some indexes (e.g. 2007) if we use older weights, we obtain a bigger bias; but for others (e.g. 2004) the bias decreases more when we are using old weights. Observing the 2008 index, we notice that the worst performance is obtained using weights that are 4 years old, but we also notice that the bias of the index is reduced if we use 1, 2 or 3 years old weights, instead of using current weights. On the contrary, the most precise value for the 2009 index is obtained using 5 years old weights. The graph 3.2-2 shows the results obtained with the SPS selection scheme. At an average level (thick line) the bias is bigger than the one obtained with SRS (note the different scales of the two graphs). This is because the units are selected with a probability proportional to size and so the system of weights gives more importance to the bigger units that are usually characterized by bigger values of the index (Toninelli, 2008). What is in common within the two sampling methods is that we obtain a more biased average path when we use older weights. But, again, this is the result of compensations of the completely different paths of the various indexes. For some of them (e.g. 2006, 2008, and 2009) the bias is extremely reduced using current weights, while for some others (e.g. 2004 and 2005) the bias increases with older weights; other indexes (like the 2007's one) have more irregular paths. Seeing this, it is clear that, using different sampling schemes, we obtain very irregular paths of the indexes that often have not much in common with each other. For this reason, it is not possible to define a precise general consequence on the index's estimates, while weights are becoming dated. This is probably because each index considers a changing systems of weights that is referred to various economical sectors; each one of them can influence, more or less, the general structure and the distribution of weights and, then, the index estimates and their bias. In the next par. 3.3 the sampling selection schemes will be compared in term of bias and standard deviation.

3.3 Comparing sampling schemes

In the following analysis we summarize by year the effect of using current rather than dated weights. A bootstrap-like selection methodology was used, selecting 5,000 samples of 1,000 units each. The evaluation of the estimates, for each considered year, is made in terms of bias and standard deviation: they give us an idea of the precision of the estimates. The bias is obtained comparing the computed simulated estimates of the index (an average of the estimates given by the 5,000 selected samples) with the corresponding actual index. The standard deviation is computed for each index obtained through the 5,000 bootstrap-selected samples. The average of these sample values is then compared to the standard deviation of the actual index. Using these criteria we evaluated, by sampling selection scheme, the percentage reduction of the bias and of the standard deviation that we obtain using current weights (tables 3.3-1/2). In the tables the cells (year/sampling scheme) where the updated weights giving a better performance are underlined in grey. The percentage reduction of the bias using updated weights is shown in table 3.3-1 (left). SRS performs better, using updated sampling weights, in two cases out of six (2005: -27.3%; 2008: -39.7%); in a third case (2007: -0,8%) using updated rather than not-updated weights we obtain round the same level of bias; in the other cases (2004, 2006 and 2009) we obtain a bigger bias (more than +47%) using updated weights; the same happens at an average level (average of the six years: +28.4%). PPS and SPS give better results using updated sampling weights in the same years (2006, 2008 and 2009), that is, in three cases out of six. Also, on average the bias is reduced by, respectively, 10.7% and 6.2%, in comparison to the bias obtained using not updated weights.

Table 3.3-1/2
Updated Vs not-updated weights estimates (bias and standard deviation percentage reduction)

	BIAS		
Index (year)	SRS	PPS	SPS
2004	56.4	5.8	17.3
2005	-39.7	17.7	27.8
2006	47.5	-43.1	-29.1
2007	-0.8	29.5	30.9
2008	-27.3	-32.6	-28.6
2009	76.3	-65.3	-83.2
Average	*28.4*	*-10.7*	*-6.2*

	ST. DEV.		
Index (year)	SRS	PPS	SPS
2004	0.2	-10.8	-15.7
2005	-3.0	-25.1	-25.9
2006	-0.5	-12.3	-17.9
2007	5.1	-14.8	-16.1
2008	-8.0	-29.3	-44.0
2009	1.5	43.8	49.1
Average	*0.1*	*-12.6*	*-14.2*

The percentage reduction in the standard deviation (in comparison with the actual index's one) using updated sampling weights rather than not updated weights is shown in table 3.3-2 (right). With SRS we usually don't get considerably better results when we use updated weights: on average, we obtain a variability that is almost the same. In fact, the difference between the standard deviation obtained with updated rather than not updated weights is reduced to 0.1% only. Anyway, for the six considered indexes (2004 to 2009) we obtain a slightly smaller standard deviation, using updated weights, in three cases on six (2005, 2006 and 2008) while for other years the standard deviation is smaller with not-updated weights. On the other hand, with the probability proportional to size selection schemes (PPS and SPS) we obtain better results in most of the cases (five out of six years) if we are using updated weights. Moreover, the standard deviation level is strongly reduced (from about -10% to -44%); over the six years PPS leads to an average reduction of -12.6%, and SPS performs even better, resulting in an average reduction of -14.2%. For 2009 only, the standard deviation increases considerably using updated rather than not updated weights (PPS: +43.8%; SPS: +49.1%).

3.3 Final remarks

Studying the variability of the weights over time, we found that sampling weights vary more in the long term, while economic weights vary more on an annual basis. Moreover, the over/underestimation of the actual index and its variance are highly different year by year: it is not possible, on average, to identify a clear trend in the evolution of the weights. Further research is needed as soon as new BR data become available. Moreover, a deeper study, taking into consideration the different economical sectors, could be useful to identify paths that may escape detection when studied globally. This will allow us not only to understand if we can define a more regular evolution of the weights of a certain economic area over time, but it will also help us to forecast their future change. To understand how often weights should be updated, we compared the values of the actual index (computed with current weights) and the values of the simulated indexes (computed with not-current weights). We also studied the path of indexes while weights are becoming dated. Considering the SRS results, we noticed that, even if at a single level the paths of the indexes are really different from one another, on average the bias of the index (and its variance) is usually reduced using weights that are no older than 2 years. For proportional to size sampling schemes, we usually obtain a bigger bias (seen the bigger importance given to the biggest units indexes) and it slightly increases too, while weights become dated. But, again, this is an average result of completely different paths. We then compared some sampling selection schemes in terms of the percentage reduction in the bias and standard deviation obtained using updated/current weights. If compared to SRS, the probability proportional to size sampling schemes give us better results in terms of bias (3 cases out of 6 and on average). Moreover, when we are using updated weights, the standard deviation of the index is reduced too (5 cases out of 6 and on average), and this means that the index is more precise. On the contrary, SRS method does not usually work better with the current weights: it gives better results more "randomly". So, we should always use updated weights. But the main issue in choosing a weights' system is that current weights are usually not available for a while, so how often should the weights be updated? As often as possible, of course. Ideally, they should be updated annually (or as soon as new data are available); in fact, sampling weights older than two years are not recommended (see graphs 3.2-1/2). If weights are updated less frequently, one could sometimes be lucky, using a system of weights that can be similar to the current one. This is because the weights do not follow a linear trend. But it is not possible to identify the system of weights more similar to a certain year (i.e. the current year) if we do not know the corresponding weights. Furthermore, the variance of the index is higher if weights are not updated resulting in an index that is less precise. We focused the second part of our work on the change in the sampling weights, in particular, so further studies are needed to measure the impact of economic weights on index estimates and the combined impact of both sampling and economic weights, eventually taking into consideration the different economic areas.

References

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 4. http://www.ssa.gov/history/reports/boskinrpt.html.

Consumer Price Index reference paper (1995). Statistics Canada, Price Division. Ottawa, Canada: Minister of Industry.

Laspeyres, E. (1884). Hamburger Warenprise 1851-1860 un die californisch-australicher soldent dukung seit 1848, Ein beitrag zur dehre von des Geldentwertung, in *Jarbüucher fur Nationalöokonomie.*

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

Producer Price Index Manual - Theory and Practice (2004). Washington, D.C.: International Monetary Fund (http://www.imf.org).

Toninelli, D. (2008). Survey Techniques: an application to prices data for the computation of price indexes, unpublished PhD thesis, Bergamo, Italy: University of Bergamo.