Enhancing the Quality of Price Indexes – A Sampling Perspective

Jack Lothian[1] and Zdenek Patak[2]
Statistics Canada[1]
Statistics Canada[2]

Abstract

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. In Canada and the U.S., and certainly in other countries, there have been questions raised about the conceptual underpinnings, the definitions and the manner, in which the data are collected and processed. 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 using this opportunity to investigate issues such as improving the frame and sample coverage, introducing two-stage random sampling techniques, improving outlier detection techniques and introducing estimation of sampling variance.

This paper sketches out our progress to date with emphasis on the sampling design, in particular, a comparison of PPS (probability proportional to size) versus SRS (simple random sampling). The results of the simulation study favour PPS but further research is required to gain a more thorough understanding of how this advantage is affected by diminishing sample size, non-response, and outlier treatment. Hence, the paper also outlines some possibilities for future development.

Key Words: Price Index, Simulation Study, Probability Sampling

1. Introduction

One of the mandates of a central statistical agency is to collect information for the compilation of the national accounts, which in turn is used to measure changes in the well-being of the national economy. To achieve this goal, a number of business surveys gather data on economic inputs and outputs at current dollar values. To evaluate real changes in the state of the national economy, one must eliminate the effects of inflation by converting the current dollar values to constant (deflated) dollars based upon a standard base year's prices. Price indexes are a tool that is used to convert all inputs and outputs collected in current dollars to constant dollars for a selected point in time.

A Services Producer Price Index (SPPI) measures the rate of change in the prices of specific services bought and sold by producers within a specific industry group. In theory, thousands of these indexes might exist and it is the national accountants' responsibility to apply them to the information derived from business surveys. In practice, the national accountants reduce the complexity of the task by assuming that a specific index is representative of a group of commodities within a group of industries. Defining the grouping procedure is a complex task that combines knowledge of data availability, conformity with international standards and the national accountants' judgment.

Collecting data for a specific SPPI is an important matter. In practice, SPPIs adopt a two-stage sampling procedure that samples representative products from a selected sample of representative establishments within an industry group. The price changes for these sampled products are then collected for a selected time period and a weighted average for the producer group is computed. It is assumed that out of the millions of transactions taking place during the reference period these selected transactions will represent the average price changes experienced by producers within the industry group.

Once frame and sampling strategies have been established, one must define formulas for computing price changes at the sample unit level and at the aggregate group levels. Because price changes are ratios, the definition of the aggregate ratio is not straightforward. The relationship between the unit ratios and the aggregate ratios is non-linear, and mathematical operations are often non-transitive and non-commutative. Also, different formulas can have different economic properties and interpretations. In addition to the above issues, there is much debate going on concerning the how to incorporate information into SPPIs on product creation, product obsolescence, and the effect of changes in quality of the product. Overall, estimating SPPIs is a challenging task.

Since its inception in the late 1800s, price index estimation has been dominated by judgmental and heuristic methodologies. Starting in the mid-1900s, scientific methodologies such as sampling started to enter the field and most central agencies continue to move away from judgmental strategies. In the last 50 years, major strides have been made in introducing standard business survey methodologies into price index estimation.

The 1980s and 1990s were a time of great turmoil in the field, particularly for the consumer price index (CPI). Throughout the world, renowned economists were criticizing the foundation and methodological underpinnings of the current methods for calculating price indices. In response to these criticisms, technical experts from around the world were brought together to discuss these issues and offer new directions.

As a result of these discussions, many national statistical agencies are revitalizing their price index estimation strategies. New indexes are being introduced, new data sources are being developed to enhance the coverage and scientific statistical methodologies are being augmented. As survey methodologists, the area that concerns us the most is improving the quality of our frames, increasing the use of scientific probability sampling strategies, and introducing quality measures.

By means of a simulation study, this paper compares two probability sampling strategies, stratified simple random sampling without replacement (SSRSWOR) and probability proportional to size (PPS) sampling, commonly used in practice in the context of estimating a price index. The study was done to answer the question of whether SSRSWOR presents a viable alternative to PPS sampling, which is the dominant probability sampling strategy used for price index estimation.

In the next section, some of the advantages and disadvantages of SSRSWOR versus PPS sampling will be discussed. In Section 3, the survey used for building the simulation population will be described. Sections 4 and 5 present the manner, in which the simulation population was constructed and the simulation study was carried out. Some observations and concluding remarks are offered in Section 6.

2. SSRSWOR or PPS Sampling?

Sampling strategies can vary markedly with systematic, simple random, PPS, judgmental and purposive sampling all being used within the same agency. In addition, each of these strategies could be used independently in the two stages, plus at times the first or second stage may be deleted. The frames used by these sampling strategies can vary strikingly as well. This is particularly true for the sampling of commodities at the second stage where judgment is frequently used to select commodities.

The current redesign of prices programs around the world faces many challenges worthy of several papers and it is difficult to do justice to the subject in a single paper. The area that we want to focus on is which sampling strategy is the most appropriate for price index estimation, SSRSWOR or PPS sampling (Ohlsson, 1998). Each strategy has its merits. SSRSWOR is easy to understand and thus it is less of a black box. In addition, there is a wealth of classical survey literature documenting how one should implement a two-stage sampling procedure and how to calculate the variances associated with this strategy.

PPS sampling's main advantage is that it is "self-weighting". Economic theory states that the change in the measured price must be economically weighted by the firm's or the commodity's contribution to industry output. The PPS sampling weights are typically the inverse of these economic weights and thus these two weights cancel. This implies that the point estimate is a simple average, and that the estimation formulas are decomposable at all levels of aggregation. Another advantage of PPS sampling is that it reduces significantly the response burden imposed on small firms.

The complexity of price index formulas and the fact that historically most samples have been judgmental in nature have not been conducive to computing variances of price indices. Quality indicators associated with estimates based on judgmental samples are more likely to follow simple *rules of thumb* rather than a theoretically sound procedure. In the absence of solid quality measures, judging the precision of an index becomes elusive. Moving away from non-probability sampling strategies is the first step on the way to being able to assess the quality of price indices in a scientific manner.

Comparing the properties of SSRSWOR and PPS sampling methods of selecting random samples for estimating a price index has the added benefit of allowing the methodologist to evaluate sample size versus precision considerations. There are many low level indices currently in production that are based on a sample of 10 or fewer (this is typical for price index samples used by statistical agencies around the world) unaffiliated price quotes at the end of the production cycle. This is often deemed sufficient to compute an index of publishable quality.

3. Wholesale SPPI

The first in the series of new SPPI surveys that Statistics Canada is developing is the Wholesale services index. It was started two years ago and has recently entered its sixth cycle. The target population consists of all businesses operating in Canada that have at least one establishment within their structure coded to one of the in-scope North American Industry Classification System (NAICS). To reduce response burden and to restrict units with a negligible impact on index calculation from entering the sample, the establishments with the smallest revenue that contribute a maximum of 5% to the population revenue were eliminated from the target population. This reduced the target population to a survey population of approximately 35,000 wholesale establishments.

The survey population was stratified by trade group (a collection of 3- and 4-digit NAICS codes used by the Monthly Wholesale Trade Survey) and by size based on estimated annual revenue. The initial sample size was driven by budgetary and resource considerations as per the recommendation of the Producer Price Index Manual, International Monetary Fund (2004). Once sufficient data have been collected, the sample may be re-allocated to strengthen areas with high sampling variance.

The sample size was set at 3,000 units, and was allocated to 17 trade groups according to an x-proportional allocation (Bankier, 1988). Subsequently, the units were allocated to three size strata, (a) certainty, (b) large take-some, and (c) small take-some. In each take-some stratum, a sample was selected using PPS sampling, giving larger units higher probability of selection.

Each participating establishment was asked to provide pricing information on three items that generate the most annual revenue. It is believed that the variable of interest, gross margin, is highly correlated with revenue. To reduce response burden, monthly data are collected on a quarterly basis. To ensure that the collected information does not contain grossly anomalous values, basic outlier detection is performed. At the end of collection, there are three – a triplet – complete and validated observations for each time period for each respondent. These observations form the basis for the frame used in the simulation study. The specific details of frame creation follow in the next section.

4. A Note On Frame Creation And Sampling

To create the sampling frame for the simulation study, all complete triplets collected in the 15-month period from October 2005 to December 2006 were extracted from the Wholesale SPPI database. Each triplet had to pass certain criteria to be deemed suitable for the simulation study. The current and previous input and output prices and gross margins (output price – input price) had to be valid and greater than zero. The ratios of current to previous input and output prices (price relatives) had to fall in the interval [1/3, 3]. This last step was performed to eliminate extreme suspect data points.

All the observations passing the above criteria were pooled into one aggregate data set that represented the preliminary simulation frame. The size distribution on the preliminary frame was adjusted to more closely match the size distributions that were observed in the actual firms on the survey frame. This resulted in many small units being cloned several times as the PPS sampling scheme that was used to draw the sample favours the selection of larger units. The resulting simulation population contained some 35,000 establishments.

Now, suppose that, based on sample (s) of y-values, we want to estimate a population (U) weighted mean $\bar{Y} = \sum_U a_i y_i / \sum_U a_i = \sum_U A_i y_i = I_L$, which is a definition of the Laspeyres Index, with $A_i = a_i / \sum_U a_i$ being the relative weight of unit i. If π_i is the probability that a unit i is in the sample, then a sample based estimate of I_L is $\hat{I} = \sum_s \pi_i^{-1} a_i y_i / \sum_s \pi_i^{-1} a_i$.

In the context of estimating a price index, the observed values y_i are elemental price indices computed at the unit level, and the relative weights a_i represent the relative importance of a unit i in the aggregate formula. Typically, a_i is a fixed measure of size – revenue (or employment, as is the case for other statistical agencies such as the U.S. Bureau of Labor Statistics) – available for all units in the population.

The sample based weighted mean can be seen as a ratio estimator, $\hat{I} = \sum_s \pi_i^{-1} z_i / \sum_s \pi_i^{-1} a_i = \hat{Z} / \hat{A}$, where $z_i = a_i y_i$ and $\hat{I}$ denotes a price index. Now, we can appeal to classical estimation theory to proceed with variance estimation, be it in the context of SSRSWOR or PPS sampling. Information on how to estimate the variance of a ratio estimator in the context of SSRSWOR and PPS (Poisson) can be found in Särndal *et al.* (1992).

5. Simulation Study

The population from which samples were drawn was based on the sample currently used for the Wholesale SPPI. An advantage of using real data is being able to assess the impact of various types of non-response on the estimated variances (current study will be extended to compare stratified SRSWOR and PPS sampling under various types of non-response, and shrinking sample sizes).

A total of 5,000 samples, each comprising 3,000 units, were generated for each sampling scheme. For PPS, in particular the Poisson variant, a unit is selected with a probability proportional to revenue (or another size measure available universally), i.e., $\pi_i = nX_i / \sum_i X_i$, where π_i is the probability of selecting a unit i, X_i is the associated size measure, and n is the desired sample size. If $nX_i / \sum_i X_i \geq 1$, the unit is selected with certainty and the corresponding total size measure is adjusted accordingly. This process is repeated until all certainty units have been identified.

For SSRSWOR, the Lavallée-Hidiroglou (1988) algorithm was used to identify optimal stratum boundaries. The survey frame was divided into one certainty and two probability strata. The desired sample size was partitioned using Neyman allocation. Within each non-certainty stratum, units were selected without replacement using the same probability of selection.

To assess the properties of each sampling scheme in terms of precision of the resulting estimates, both empirical sampling bias ($\hat{B}$) and variance ($\hat{V}$) were computed. They were defined as

$$\hat{B} = \frac{1}{5,000} \sum_{i=1}^{5,000} (\hat{I}_i - I_L)$$

and

$$\hat{V} = \frac{1}{5,000} \sum_{i=1}^{5,000} (\hat{I}_i - \bar{\hat{I}})^2 ,$$

where $\bar{\hat{I}} = \frac{1}{5,000} \sum_i \hat{I}_i$. To assess the properties of an estimator, one typically computes the relative empirical sampling bias, but in the context of estimating a price index, dividing a very small change by a number close to one does not change the results.

The parameter I_L is the true population Laspeyres price index formulated as follows

$$I_L = \frac{\sum_U P_{ck} Q_{bk}}{\sum_U P_{bk} Q_{bk}} = \sum_U A_{bk} \frac{P_{ck}}{P_{bk}} = \sum_U A_{bk} y_k ,$$

where the subscripts b and c denote the base and current periods, P denotes price, Q is quantity, and A_{bk} is referred to as the base period economic weight – relative contribution of an entity (business unit or item) to the overall index computed at some level of aggregation (industry, province, etc.) – defined as $P_{bk} Q_{bk} / \sum_U P_{bk} Q_{bk}$. The population index I_L is the quantity that most statistical agencies estimate in practice.

In economic literature, one can find many different index formulas that have been defined over the years to estimate the true price movement while still being operationally feasible and deliverable in a timely fashion. The interested reader may consult the Producer Price Index Manual, International Monetary Fund (2004) for a brief overview of the most commonly used in practice indices.

We confine the simulation study to the geometric mean index (Jevons Index) and the arithmetic mean index (Laspeyres Index) as both have a number of desirable properties, not the least of which is easy implementation. Economic theory tells us that the Laspeyres Index is typically upward biased and hence it provides an upper bound on the true index. The geometric mean is always lower than the arithmetic mean and as such, it is closer to the true index.

The two tables below show the empirical sampling bias and standard deviation corresponding to the two indices in the context of SSRSWOR and PPS sampling. It should be noted that the bias associated with the Jevons Index is economic as both sampling methods are unbiased. The overall bias is negative, which only confirms the relationship between the arithmetic and geometric means. Both sampling methods produce similar levels of precision with PPS sampling marginally outperforming SSRSWOR. However, this gain in performance is not statistically significant.

In Table 2. the economic bias disappears as the definition of the variable of interest coincides with the population parameter. Once again, PPS sampling outperforms SSRSWOR with precision results being almost identical to those reported in Table 1.

The simulation study was designed to measure monthly price movement and although the differences between the two sampling schemes are small, they may become significant when annualized. Assuming that monthly changes are relatively constant over time, the annual rate of change estimated by SSRSWOR is 0.8% compared to 2.3% for PPS sampling when using the Jevons Index. The annual rates are obtained by first correcting for the bias, and then multiplying the difference between one and the estimated parameters by twelve.

Both numbers seem plausible as estimates of annual price movement and need to be compared to the "true" price index, which will be investigated shortly.

Table 1. Geometric mean (Jevons Index) at unit level

Trade Group	Bias (SRS)	Std Dev (SRS)	Bias (PPS)	Std Dev (PPS)
A	0.006	0.009	0.006	0.008
B	-0.024	0.004	-0.024	0.004
C	-0.004	0.005	-0.004	0.001
D	0.017	0.006	0.018	0.001
E	-0.003	0.004	-0.002	0.004
F	0.009	0.005	0.009	0.004
G	0.017	0.011	0.016	0.010
H	0.001	0.004	0.001	0.004
I	-0.092	0.008	-0.093	0.003
J	0.007	0.002	0.007	0.001
Overall	-0.007	0.002	-0.006	0.001

Table 2. Arithmetic mean (~ Laspeyres) at unit level

Trade Group	Bias (SRS)	Std Dev (SRS)	Bias (PPS)	Std Dev (PPS)
A	0.016	0.010	0.015	0.010
B	-0.016	0.004	-0.016	0.004
C	-0.001	0.005	-0.001	0.001
D	0.021	0.006	0.021	0.001
E	0.003	0.004	0.003	0.004
F	0.020	0.005	0.021	0.004
G	0.042	0.012	0.040	0.011
H	0.007	0.004	0.007	0.005
I	-0.072	0.009	-0.073	0.004
J	0.017	0.002	0.017	0.002
Overall	0.001	0.002	0.000	0.001

Several other index formulas were studied in addition to the two exposed in this paper. Apart from the economic bias showing considerable range eliminating some index formulations as not being viable, the statistical properties under both designs were comparable.

The simulation results did not change appreciably when gross margin, the true variable of interest that would not be available outside of a simulation study, was used as the size measure for sample selection and economic weighting. This is most likely due to a large overall sample size.

6. Concluding Remarks

The two tables show very little difference in the statistical properties of a price index based on either stratified SRSWOR or PPS sampling. This is true in the ideal setting of the current simulation study. To inject a touch of reality into the simulation study, we would like to investigate what happens when real life phenomena are incorporated such as imperfect size measure, other types of misclassification, non-response, outliers that may go undetected, and various types of imputation for missing data.

We would also like to modify the population file to allow for the computation of a "true" price index to verify the conjecture that the results reported for Laspeyres Index also hold for the "true" price index, i.e., statistically no difference between PPS and SRS.

It is not clear how sample size would affect the results. To assess how robust the two designs are against diminishing samples, the simulation study needs to be extended to cover sample sizes that are typically seen in real price index surveys. The sample sizes may be reduced even further to identify the "breaking point" of each design.

It should be noted that De Haan (1998) and Dorfman (2006) produced simulations using a limited set of scanner data that indicate that a simple cut-off strategy performs as well as SRS. We also hope to explore this option using our dataset.

Acknowledgements

The authors would like to thank the following for their comments and suggestions, which contributed to improve greatly the final version of the paper: Fred Barzyk, Jean-François Beaumont, George Beelen, Sylvie Gauthier, Pierre Lavallée, and Wesley Yung. The views expressed in the paper are those of the authors and do not necessarily reflect the official position of Statistics Canada. All remaining errors are those of the authors.

References

Bankier, M (1988), *Power allocations: determining sample sizes for subnational areas.* The American Statistician, **42**, 174-177.

Boskin M. J., Dulberger E. R., Gordon R. J., Grilliches Z. and Jorgenson D. (1996), *The Boskin Commission Report: Toward A More Accurate Measure Of The Cost Of Living*.

De Haan, J., Opperdoes, E., Schut, C. M. (1998), *Item Selection in the Consumer Price Index: Cut-off Versus Probability Sampling*. Survey Methodology, **1**, 31-41.

Dorfman, A. H., *et al*. (2006), *On Sample Survey Designs for Consumer Price Indices*. Survey Methodology, **2**, 197-216.

International Monetary Fund, International Labour Organization. (2004), *Producer Price Index Manual: Theory and Practice,* Washington, DC: International Monetary Fund.

Lavallée, P. and Hidiroglou, M. A. (1988), *On the Stratification of skewed populations*. Survey Methodology, **14**, 33-43.

Ohlsson, E. (1998), *Sequential Poisson Sampling*. Journal of Official Statistics, **14**, 149-162.

Särndal, C.-E., Swensson, B., Wretman, J. (1992), *Model Assisted Survey Sampling.* New York: Springer-Verlag.