Integrating Culture Industries into Canada's National Economic Surveys Program

Mary March[1], Paddison Wong[2]
Statistics Canada[1]
Statistics Canada [2]

Abstract:

Statistics Canada's cultural annual surveys are being integrated into the Agency's economic data program beginning with the 2004 reference year. The project has faced numerous challenges including a requirement to develop a backcasting methodology (using data collected by the previous program, administrative data and data collected by the newly integrated survey) to provide a historical data series corresponding to the changed target population. This paper discusses some of the challenges that were faced and will explain and evaluate the methodological solutions that were developed for the backcasting problem.

Keywords: Integration, Backcasting, Annual Survey,

1. Introduction

Over the last decade Statistics Canada has developed an integrated economic statistics annual survey program which covers a wide range of industries including services industries. The Business Register (BR), a central frame kept up-to-date with information from administrative sources and survey feedback, is at the core of this program. Concepts and questionnaires and survey methods used by the program have been standardized. Frame processes and input sources have progressively been improved. Data are compiled, analyzed and made available under the framework of the North American Industry Classification System (NAICS).

For a number of years, the agency measured cultural activities within a culture statistics program which existed separately from the economic program. This cultural statistics program's concepts and methods evolved independently from the economic framework.

Some aggregate financial information for the Cultural industries was produced within the economic surveys program by Services Industries Division. According to these data (for reference year 2002), the cultural industries were accounting for 3.8% of Canada's GDP (Gross Domestic Product) and providing employment for an estimated 600,000 persons.

It has become a struggle to manage the cultural statistics program which has been faced with increasing financial pressures and additional requirements for information. One impact of these pressures has been a decrease in the frequency of these surveys.

A streamlining initiative was proposed – to integrate the culture statistics program with the economic surveys framework. This proposal was an attractive one because integration would reduce overall costs while at the same time improving coverage of the economy and data quality (frequency, usability, coverage and coherence) of the information on culture being produced. It would also extend coverage of the culture statistics program to include cultural activity in the "for-profit" sector.

A project was initiated and a joint team representing the economic and culture programs was charged with the goal of integrating the culture surveys with the economic program over two reference years (2004 and 2005).

1.1 Business Register (BR) reconciliation

Since the culture surveys would now use the centrally maintained BR as a frame, it was important to reconcile the register with the frames previously maintained and used by the culture program for these surveys. Reconciliation consisted of matching the frames and then resolving previously surveyed units that were either missing from the register, on the register but recognized as belonging to another industry category or on the register but included as part of another larger unit.

1.2 Developing a sampling strategy

The culture surveys had relied on a census of their populations. With the change to a larger target population, a full census was no longer feasible. Therefore, it was decided to sample units with probability of selection determined by a unit's revenues. This is the approach used by economic surveys

1.3 Adapting to generic economic survey concepts, questionnaire design and survey processing methods and analysis

New more consistent survey questions were developed and tested and processing systems were adapted to handle them. Finally, the analysis was adapted to the new environment including pre-release approaches to data validation.

1.4 Enabling users to contend with the change

Policy makers and analysts had had a long tradition of relying on historical data series from Statistics Canada's culture surveys. Changes in content and coverage of the information that would now be provided would significantly affect data series available to these survey users. Of particular concern was the expected change in coverage of the estimates which would mean breaks in the data time series.

Statistics Canada agreed to "**backcast**" the first culture estimates obtained within the economic program to provide estimated historical series for key variables corresponding to the new target population.

The authors of this paper were responsible for developing and implementing the backcasting methodology. A description of the process by which these backcasted estimates were obtained, some data quality measures and an evaluation of the results are contained in the remainder of this paper.

2. Culture industry series to be backcasted

The culture surveys program yielded series covering production data and other activities for the following list of industries:

- performing arts,
- heritage,
- book publishing,
- periodicals publishing,
- film production,
- film post-production,
- film distribution,
- sound recording and
- movie theatres.

In the heritage and performing arts industries, coverage was focussed only on the not-for-profit sector.

In planning the integration project it was decided that the surveys of the heritage, performing arts, and book publishing industries would be integrated as of the 2004 data reference year. The remaining surveys (periodicals publishing, sound recording, film production, film post-production, film distribution and movie theatres) would be integrated as of the following reference year, 2005.

The remainder of this paper focuses on backcasting for the three surveys integrated as of the 2004 reference year.

3. Understanding changes in coverage due to integration

3.1 Differences due to use of the BR

From the beginning, the change in survey frames (which are derived from different sources) was expected to affect the survey's coverage.

In the initial reconciliation matches between the BR, which contains classifications by NAICS (North America Industry Coding System) codes, and the previous culture survey files, a number of new units were found that had not been part of the previous survey population. Subsequently, however, after the collection operation was completed, many of these units were found to be out-of-scope to the survey, either because of an incorrect industry code, or because they were no longer in business. This was to be expected because it is a common occurrence with any survey using the BR as a frame for the first time. When records are initially added to the register, industry codes are imputed by automated scanning of text activity descriptions on the input administrative records. Many thousands of records are added each year and resources are not available to verify the classifications. Also, there is a significant lag time before the register's processing systems determine that a business on the register has died since administrative or survey feedback are relied on to confirm deaths.

At the planning stages of the first three surveys, the eventual out-of-scope rates to be expected for new units in the BR culture population that had not previously been surveyed were difficult to predict. In our backcasting strategies, we needed to be prepared for a potential significant increase in the in-scope population sizes for the culture surveys. Actual out-of-scope rates for BR units are shown in Table 1 below.

Table 1

2004 Survey Out-of-Scope/Dead Rates	
Book Publishing	27.3%
Performing Arts	28.0%
Heritage	24.6%

3.2 Differences due to adding "for-profits"

The previous culture statistics program surveys of heritage and performing arts industries had excluded "for-profit" organizations. With integration, the for-profit sector was added to these culture surveys within the economic program with a plan to provide separate domain estimates for the for-profit and non-profit parts of these industries. A derived variable on the BR was used to classify units as belonging to each sector.

4. The sampling strategy

The integrated culture surveys adopted a standard business survey sampling approach (as described in Business Survey Methods, Cox et al 1999 and in the following paragraph) which takes advantage of the skewed distributions of business financial variables found in most industries.

BR variables used for defining planned estimation domains were geographic (provincial) codes, a for-profit/not-for-profit indicator and a foreign/domestic ownership indicator. An algorithm (Hidiroglou-Lavallée) used the revenue variable to create four sub-strata, each with an appropriate sampling rate – the largest units were selected with certainty (take-alls); the next two size groups were subject to two levels of sampling probabilities; and the last group, accounting for 5 % of the revenue, were not surveyed. Standard practice in economic surveys is to reduce respondent burden by not surveying the smallest units that contribute to the bottom 5-10% of an industry's revenue and to use data from tax sources to account for them in final estimates. A 5% cut-off was chosen for the culture surveys.

Previous culture surveys had been complete censuses of units on the survey frame regardless of their size.
For analysts and data users, sampling errors including variance and design CVs were a new reality. Some previously asked questions, relevant to but a small proportion of the population, were clearly no longer worth asking unless the population they pertained to a well-known subset of the take-all portion of the survey enabling enough responses to obtain reasonably good data . A *must-take* category of units belonging to important small subsets of the population was defined. Subject matter officers could designate units to this category even though revenue measures alone might not indicate that they were important enough for a census. In addition, the "take-all" algorithm was applied to key variables other than revenue on the previous survey database to propose other potential must-take units.

Sample sizes were inflated across the board under the expectation of a 30% out-of-scope rate for the BR units.

5. Backcasting

Data users requested that we "backcast" the integrated survey estimates for the two most previous reference years for each survey. The backcasting reference years varied depending on the survey and are shown in Table 2 which follows.

Table 2

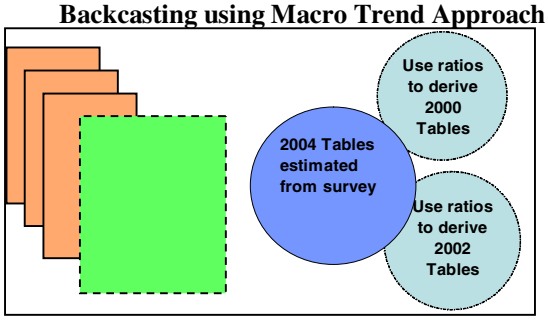
Industries	Previous survey years
Performing Arts	2001,2003
Heritage institutions	1999,2002
Book Publishers	1998, 2000

5.1 Trend Approach.

For variables that existed in both versions of a survey, its historical culture program version and the new integrated version, we could have used 2004 survey data files to produce the same tables as were published in the previous two reference years but for the 2004 target population. At the same time, the data for the two previous years could have been used to forecast estimates for 2004 for the previous target population. For the corresponding table cells, trends between previous years and the forecasted 2004 tables could have been computed and used to adjust or "backcast" the 2004 tables for the full target population to the previous reference years.

This method is illustrated in Figure 1. below.

Figure 1.
Backcasting using Macro Trend Approach



This approach was attractive to some of our colleagues since it would be simple to implement and explain. There is some inherent bias associated with imputing the trend of the previous population to a differently defined target population. (We especially had reservations about imputing trends from the not-for-profit sector to the for-profit sector.) Nevertheless,

since the method was not difficult, we agreed to compute some backcasted estimates using this approach and compare the results with those we obtained by the method we did use as an evaluation.

5.2 Longitudinal survey approach

Data were available from three different sources for the culture surveys population for the two previous reference periods.

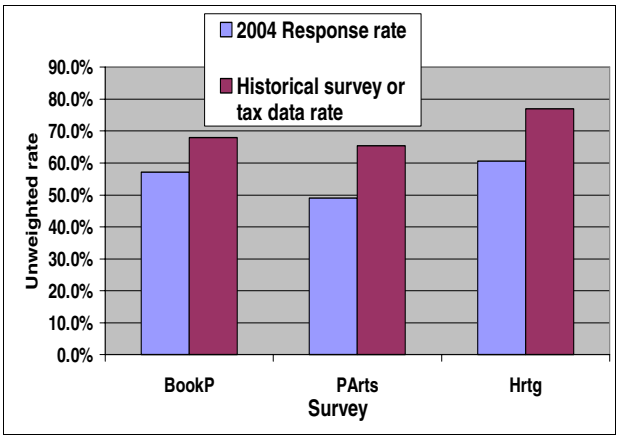
- Data collected by the culture survey before integration for units in the previous target population and using a different questionnaire.
- Tax data (from income and sales tax sources). These records could be at a level other than the establishment level used in these surveys.)
- For a limited number of units, data had been collected by an arts, entertainment and recreation survey which had been part of the economic statistics program. It was likely that some of the units with larger revenues in the performing arts industry (for-profit and not-for-profit) had been selected for that survey with certainty.

We decided to match our 2004 samples with data files from these three sources to assess their comparative coverage. We realized that the 2004 sample selected from the business register could under-cover units that had been alive in previous reference periods but no longer existed and could have been removed from the business register. (However, the time lag in frame updates would certainly be to our advantage for preserving these units on the business register.)

All-in-all we found we could match about 70% of the 2004 sample which is slightly better than the response rates we obtained for culture surveys in 2004 data collection.

2004 response rates are compared to matched rates in the most recent reference years in Figure 2 below.

Figure 2.
Survey response vs. historical match rates



We decided to create a "survey file" for each of the backcasted years. For each previous reference year we would "collect" data for the units from the historical files for the units that had not been determined to be out-of-scope in 2004 collection.

Unfortunately, the only reconciled frame available was the 2004 frame. (It was a too costly and time consuming prospect to consider trying to create a reconciled frame for the historical years.)

One potential benefit of using the same sample in estimating the backcasted years would be a reduction of variance in year to year comparisons. This is usual practice in the design of longitudinal sample surveys.

We decided to use historical counts from the BR to estimate population sizes for the target industries (realizing that we would be missing the corrections made subsequently during the reconciliation) to calibrate the survey weights for the previous reference year survey files.

Calibration was done within a NAICS. Generally speaking, for the units representing only themselves in 2004, their weights were kept the same (i.e. weight =1) in the backcasted years. Otherwise the calibrated weights were given by

$$w_b = w\frac{N_b}{N}$$

where w and N are the weight and population size above the 5% not surveyed threshold whereas w_b and N_b are the adjusted weight and estimated population size (from historical BR counts) above the not surveyed threshold and in excess of the units representing only themselves in the backcasted year. When N_b was too small or negative, we included some

of the units with weight =1 in the calibration to increase the size of N_b.

It was fairly straight-forward to set up processing for the backcasted survey files although somewhat labourious given that three 2004 processing files for three 2004 surveys became a set of nine files to be processed and validated.

Data were imputed for backcasted records that could not be matched. Imputation made use of tax data and 2004 collected data that could be adjusted by backwards trends within the same imputation groups of records. (A reversal of our usual economic survey approach where imputation uses historical data adjusted for forward trends.) The analyst working with the files carefully removed outliers in determining these trends as would also be done in processing a normal economic survey data file.

6. Backcasting results

6.1 Matching rates

With the book publishing survey, we were able to link 274 out of 332 in-scope sample units or 82%.
Weighting these units by values of the revenue variable on the 2004 frame file, the matched rate rose to 89%. 54 units accounting for 4 % of the revenue were not found in the previous sample and were imputed from tax sources.

The response rate for the in-scope units for the 2004 survey had been 79% so, all-in-all, the quality of the resulting files was quite good.

For the heritage survey, 79% of the 2004 reference year in-scope sample units were linked to 2002 survey units. Another 15% could be linked to tax information. The linked units totalled 738 out of 788 and 96% weighted by revenue.

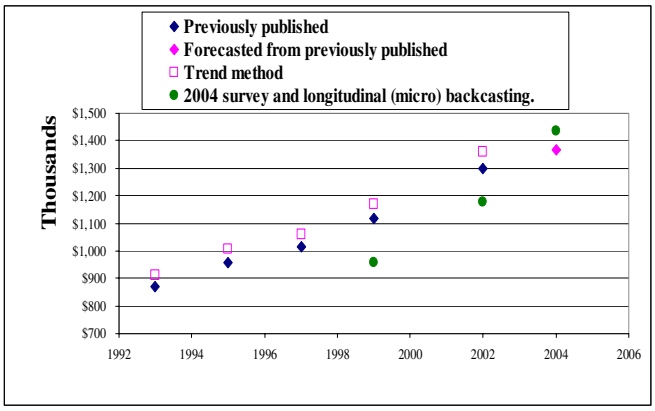
For the performing arts survey, 90.6% or 1352 out of 1491 of the 2004 in-scope units could be matched to 2003 survey or Tax data. (This was the most recent reference year in the backcasting exercise.) However there was a large difference in matching rates for the 991 in-scope for-profit units which had not been in the culture program reference population previously. 123 or 12 % of the units could not be matched and 757 or 76% could only be linked with tax data. This compares with 70 or 14% of the not-for-profit units that could only be matched to tax data.

6.2 The effect of changes in variable definition

The process mentioned in **1.3 Adapting to generic economic survey concepts, questionnaire design and survey processing methods and analysis** had a significant impact on our ability to backcast estimates for the culture surveys. Although the culture surveys had included financial variables, the variables had not necessarily been consistent with the definitions of variables used in economic surveys or in tax data programs. Sometimes the differences were subtle but they caused problems in processing the new 2004 data, where historical data needed to be used judiciously if it was an information source when imputing for nonresponse. Similarly, the backcasting process needed to checked carefully to ensure that a change in variable definition was not interpreted as a change in trend. The fact that we were working with micro data helped in analysing apparent differences.

6.3 Comparison with the Trend Method

Graph 1 compares results of the two methods applied to the heritage survey variable named Operating Revenue.



The two backcasted estimates are lower, in this case, than the previously published estimates and the estimates that would have been produced using the Trend method. This does not suggest that the backcasting underestimated. If the previous frame was not picking up all of the births, the year to year trend would be flatter and projecting it back would yield higher estimates.

Looking at similar graphs for other variables, we found that it is the 2004 estimates that determine the level of the trend estimates. Where the 2004 estimate was equal to the forecasted value based on the previous series, the backcasted values are the previous series.

The backcasted estimates were released at the same time as the 2004 estimates and were used in the validation of the data. Analysts found it useful to have a database of micro-records to support analysis of trends.

7. Conclusion

With economic programs, a change in coverage of a survey can be supported by backcasting if there is information available from administrative sources and match rates are reasonable. Sometimes, it can be done relatively cheaply if the survey infrastructure developed for the new survey can be applied to manufacture historical survey files. The resulting backcasted database can be used by analysts to validate trends that they observe. The only concern is the lack of a historical frame to provide survey weights.

The trend method has merit but underlying data is not available to analysts for validation analysis.

References

Business Survey Methods, Wiley Publications, Brenda G**.** Cox et al editors, 1995

The Canadian Heritage Website.
http://www.canadianheritage.gc.ca/index_e.cfm

The Canadian NAICS Website
http://www.statcan.ca/english/Subjects/Standard/naics/1997/naics97-intro.htm
North American Industry Classification System (NAICS) 1997 - Canada

From the Statistics Canada website.
http://www.statcan.ca/english/ads/87-004-XPB/pdf/fcis2003015001.pdf
The impact of the culture sector on the Canadian economy
by Vik Singh