Rolling Estimates and the Dynamic Assessment of Quality: a Generic Approach and a Source of Efficiencies

Etienne Saint-Pierre, Statistics Canada (Etienne.Saint-Pierre@statcan.gc.ca)

Paper prepared for the European Establishment Statistics Workshop, Nuremberg, Germany, September 9-11, 2013

Introduction

Statistics Canada is currently undertaking a major integration project for its business statistics surveys, the Integrated Business Statistics Program (IBSP). The IBSP will provide a common processing framework and common methodologies for the various business surveys conducted at Statistics Canada. From 2014 to 2017, more than 150 surveys will migrate into this new integrated and harmonized framework[1].

The three main objectives of the new integrated framework are the achievement of efficiencies, the enhancement of quality assurance and the improvement of the responsiveness in the delivery of new business statistical programs.

To meet these objectives and to build a generic framework that is flexible enough to be exportable to a large number of heterogeneous business statistical programs, every step of the business survey cycle has been reviewed, harmonized and optimized into a series of coherent processes.

This paper will focus on one of the component of this integrated processing system, namely the Rolling Estimates (Godbout et al., 2011). The Rolling Estimates is a process where estimates and quality measures are produced and analysed on an iterative basis until an acceptable level of quality is reached. This new Rolling Estimates approach will be used to dynamically manage non-response and failed edits follow-ups strategies as well as analytical activities. The key concepts and methods of the Rolling Estimates and supporting quality indicators are presented in the first section of the paper.

The implementation of this new process has been preceded by a large-scale prototype. The Rolling Estimates approach with its new set of quality indicators and score measures was simulated for nearly 50 annual business surveys in 2012. The promising results on the potential to reach quality targets earlier in the survey cycle with reduced follow-up activities will be presented in the second section of this paper.

The last section will cover the fundamental changes that the new approach induces at the different stages of the survey cycle.

1- From a Linear Model to the Rolling Estimates

1.1 Business Surveys at Statistics Canada –The Current Approach

The vast majority of the business surveys conducted at Statistics Canada use a sequential approach to processing that consists of completing each stage of the survey cycle before starting the next. First, significant efforts are devoted to collect data and validate the collected information. Some of the main priorities of this process are maximizing the weighted response rates and minimizing the outstanding cases with failed edits within a given budget and collection time frame. Once the information is collected and cleaned (through follow-up calls or by the analysts), imputation operations are done and the analysts can, for a second time, validate and clean a set of micro records. Then in a third step, the macro estimates are calculated and analysed, giving a third and final opportunity for analysts to manually correct data. At

[1]For more details on the IBSP project, see Ravindra (2012)

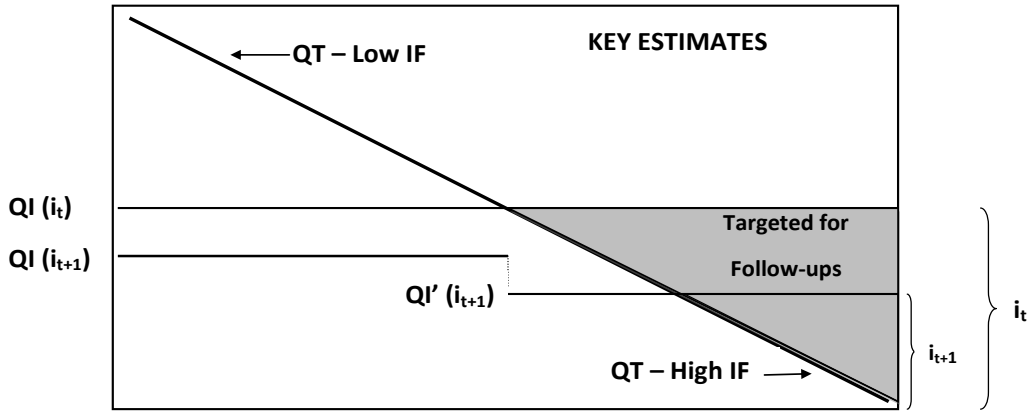
the very end of the process, quality measures and reports are produced. The imputation rates, sampling variance and quality ratings found in these reports help inform users to what extent they should apply caution in using the data. The cycle of annual surveys is spread over a period of 8 to 15 months depending on the complexity of the survey.

The implementation of the IBSP project led to an in-depth review of our modus operandi in order to reach our efficiency goals without compromising the quality of our statistics. To achieve this balancing act, the reduction of low impact manual interventions was targeted. But the cornerstone of the new model is the active management of collection and analytical activities based on the on-going assessment of the quality of the data as it becomes available. Measurement and analysis of the quality are of paramount importance in the new model.

1.2 **The Rolling Estimates and the Quality Indicators – The New Model**

Rather than starting at the micro level, the Rolling Estimates process presents the analysts with a complete set of estimates for all the survey variables. As soon as there is sufficient collected and administrative data, estimates can be produced by integrating data from multiple sources and imputing for non-response. The production of these estimates is accompanied by quality indicators for each domain of estimation and measures of impact scores for each unit in the sample. These measures become a pivotal element to identify domains and units that should be targeted in priority for follow-ups as well as to recognize domains that are of sufficient quality to be certified by analysts. The new approach for non-response and failed edits follow-ups can be summarized in the following manner: follow-ups are done only on a subset of units that can significantly contribute to improving the quality of key domains of estimation which have not reached their target quality. Figure 1 contains a graphical explanation of concepts used in the new model.[2]

Figure 1- Only units in domains of estimation where the Quality Indicator (QI) is higher than the Quality Target (QT)



[2] Methodological and mathematical details related to the quality indicators and measures of impact scores in IBSP can be found in Turmelle et al. (2012).

1) Prior to the survey cycle, a series of key variables and key domains of estimations are identified and combined to create key estimates. For example the combination of two key variables (value of manufacturing shipments and total salaries) with key domains (200 industrial groups) will give a set of 400 key estimates. For every iteration and for each of these key estimates, Quality Indicators (QI) are produced. As detailed in Turmelle et al. (2012), the Quality Indicator used in the IBSP will be a combination of these quality measures:

Sampling Coefficient of Variation & Imputation Coefficient of Variation & Pseudo Relative Bias

where the measure of pseudo bias is derived from a comparison between the predicted value derived for each key variables and the collected information. The measures of imputation CV and Pseudo Relative Bias are new in the quality assessment of business surveys at Statistics Canada.

2) Secondly, an Importance Factor (IF) is assigned to each key estimate to rank them from the most to the least important. As default rules, key estimates with the largest historical values will have the largest Importance Factor and vice versa. However the ranking of the key estimates will be amendable by the analysts after validation. For example, if an industry is expected to have a phenomenal expansion, its ranking should be readjusted to a higher level. The Importance Factors are based on previous year's value and input from the analysts.

3) Then, based on the Importance Factors, Quality Targets (QTs) are established prior to the survey cycle for each key estimate. The higher the Importance Factor for key estimates, the better the quality should be. In other words the Quality Target (the targeted CV) would be lower.

4) Finally, at every iteration of Rolling Estimates (i_t), QIs are produced for every key estimate and compared to their respective Quality Target.

As shown in Figure 1, at iteration (i_t), only units in domains of estimation for which QI(i_t) is greater than QT will be targeted for direct telephone follow-ups. All other domains would not be targeted for follow-ups (other than fax or e-mail reminders) and could be looked at by the analysts. At the subsequent rolling estimates iteration, with more collected or administrative data available, $QI_{(t+1)}$ will be lower for most domains of estimations. However, domains targeted for follow-ups after (i_t) will see their QI decline faster ($QI'_{(t+1)}$) as units having the most potential to bring the QI to the QT level would have been targeted. After the iteration (i_{t+1}) only domains for which $QI_{(t+1)}$ is greater than QT will need to be followed-up. The goal is to bring QI for all key estimates under their respective QT. Quality for non key-estimates will not be assessed in the Rolling Estimates process.

In the Rolling Estimates approach, a list of priorities for collection is built after each iteration. Only units eligible for phone calls in domain of estimation targeted for follow-ups are on this list. To minimize the number of follow-ups required to reach the Quality Target, a series of measure of impact scores (MI) are measured and assigned to each record. The value of a MI score represents the impact of a unit on the value of the QI if for this unit, values are converted from imputed to reported[3]. Units with the largest MI scores are ranked the highest on the list. In other words, units imputed based on a poor model or with large discrepancies between their predicted values and their current values will be targeted. If for a

[3] For a given survey, a business could contribute to several key estimates and quality indicators. All the QIs and MI are combined into a global score (Global QI and Global MI) that is used to manage active collection at survey level. The higher is the Global MI for a unit, the larger is the impact of this unit on the Global QI. For more details, see Mills et al. (2013)

domain targeted for follow-ups, the top N units with the largest MIs are sufficient to bring the QI to the Quality Target level, only these N units will be on the list.

2-Simulation of Rolling Estimates and Impact of the New Model on the Volume of Follow-Ups

Rolling Estimates were produced for 47 annual surveys. A wide assortment of surveys in terms of economic sectors and complexities were covered. Surveys in the distributive trades, manufacturing and logging sectors, in the services industries programs as well as a large survey measuring capital expenditures by the private and public sectors were included. Four iterations with a complete set of estimates, QIs and MIs were produced.

The simulation has been used to test various QIs and MIs on different types of variables (commodities versus financial variables for example) to determine if a more targeted collection and follow-up process would result in less units being contacted without impacting the overall quality and timeliness of the estimates.

Table 1 summarizes some key results coming from the simulations if the collection would have been managed dynamically using QIs[4].

Table 1. Summary of Results of the Rolling Estimates Simulations[5]

	Collection entities	Units Followed up for Non - Response	Units Followed up for Non - Response after 1st RE iteration	Number of Key Estimates	Percentage of Attainable Quality Target Met			Reduction -Units Followed up for Non - Response	
	1000's				Iteration 1 (June 29)	Iteration 4 (October 1st)	Using QI/MI	1000's	%
Total	47.3	27.9	14.7	8.6	76%	85%	98%	9.5	34%
Manafacturing	11.5	6.0	2.6	1.4	71%	83%	99%	1.2	20%
Distributive trade	10.1	7.3	4.2	3.6	82%	89%	97%	3.3	45%
Services Industries	25.6	14.6	7.9	3.7	72%	82%	98%	4.9	34%
Electronic questionnaires	*7.6*	*4.0*	*3.1*	*1.0*	*71%*	*82%*	*99%*	*2.0*	*49%*

At the time of the first iteration, a total of 13,200 collection units had already been followed-up for non-response over a period of four months. Data obtained from collection or auxiliary data (considered as a positive response in the calculation of quality indicators) have resulted in 76 per cent of key estimates reaching their Quality Target as defined in the new method.

A total of 14,700 collection units were followed-up for non-response between the first and last iterations. Despite these follow-ups, only 9 per cent additional key estimates reached their Quality Target using the current model. The absence of accurate indicators that take into account the quality of the imputation models and regular comparison with targets (other than the weighted response rate) in the current model led to many follow-ups on units that contributed only marginally to the overall quality improvement.

[4] In the simulation, the only quality measure used in the analysis of the results was the total coefficient of variation which includes both sampling and non-response variance. The level is reflected by the QI. It is important to note that 10 per cent of the key estimates never reached their Quality Target since the non-response portion of the CV could never be reduced enough to meet the target. In these cases, the sampling portion of the CV is above the Targeted CV.

[5] The results of the simulation presented in the table are from Mills et al. (2013)

The calculation and comparison of QIs with their respective Quality Target for every iteration helps identify which key estimates would have potentially reached their quality target should the new approach be used. The method of Rolling Estimates could have generated up to a 34 per cent reduction in the number of collection units on which non-response follow-ups would have been required while allowing 98 per cent of the 8,600 or so key estimates to reach their Quality Targets when attainable. As shown in Table 1, reductions were spread across the various programs ranging from 20 per cent to 45 per cent while allowing Quality Targets to be reach for 97 to 99% of the key estimates. The simulation clearly shows the Rolling Estimates can generate efficiencies, reduce response burden and improve the quality of the estimates.

The full potential of the new approach comes from the ability to move resources after each iteration to follow units with the highest MIs contributing to key estimates that are under the pre-determined Quality Target. If the Rolling Estimates method had been used from the first iteration, follow-ups would have been concentrated only on the 24 per cent of key estimates with QIs higher than their respective Quality Target. The old model was based on prioritizing follow-ups to maximize the response rates and did not take into account the quality of the imputation models. This explains why numerous follow-ups were made on domains that had already reached their Quality Target (according to the IBSP quality definition) and not enough on domains showing insufficient quality.

3 - Optimization of Other Processes to Support the Use of Quality Indicators in the Active Collection Management

For an optimal management of collection and analysis activities with the Rolling Estimates, various processes have been reviewed, modified and optimized.

Sampling: Small samples for domains with large Importance Factor should be avoided since the non-response portion of the CV could potentially never be reduced enough to compensate for the sampling portion of the CV. It is essential that sampling in the IBSP is coordinated with the definition of the QI/MI parameters to avoid having unattainable quality target for key estimates.

Follow-Up Strategies: All units will not be targeted for follow-ups. Only units for which the conversion from an imputed status to a reported status would have a significant impact on the QI will be targeted. Failure to convert these units into respondents will result in having to potentially follow-up several marginal non-respondents to reach the desired quality targets. This would in turn affect the possibility to generate cost reductions. The use of a collection paradata to assess the likelihood of converting non-respondents into respondents, faster escalation strategies for non-response follow-up and close monitoring of these key units by analysts with their rapid intervention to provide reliable data via external data sources as replacement for non-response are essential.

Use of Electronic Questionnaires: Under the IBSP, the primary mode of collection will be with electronic questionnaires. Claveau et al. (2012) found that significant response can be obtained without making phone calls for the first two or three months of collection by using frequent e-mail reminders. This low cost approach works but only up to a certain point at which the response rate plateaus. By delaying phone follow-ups to the point where e-mail reminders become inefficient, it gives a chance to a greater number of key estimates to reach their Quality Target before more costly phone follow-ups start. For surveys using e-questionnaires, Results in Table 1 show that with a smaller proportion of phone follow-ups conducted before the first iteration, the proportion of key estimates meeting their Quality Target was similar to other surveys. This translated into a greater potential for efficiencies (a 49 per cent reduction in the number of potential units to follow-up) than other surveys using more traditional modes of collection.

Collection Edits: Only units having an impact on the QI will be followed-up. Since the QI and the MI are only calculated on a limited set of key variables, resolution of failed edits on non-key variables will have no impact on the QI. Collection edits on non-key variables should be limited.

Response Rates: For decades, the weighted response rate was used as a target, for the monitoring of collection progress and for the prioritization of collection efforts. The new quality indicators will change this. Under the new model, domains of estimation with relatively low response rates and for which solid imputation models for non-respondent units are used could be deemed to be of sufficient quality to be closed.

Conclusion

The Rolling Estimates is a new integrated approach adopted by Statistics Canada to handle the processing and the active management of collection and analytical activities for a large number of Business surveys. The feasibility of this new approach has been successfully tested with a large-scale prototype. The results show that the Rolling Estimates model has great potential to help achieve the objectives of efficiency and quality of the IBSP by dynamically identifying a subset of non-respondent units that will contribute into the improvement of the quality of the estimates if they are converted to respondents. Furthermore the production of a complete set of estimates and series of quality indicators on a regular basis help analysts in targeting the corrections to be applied to the data by favoring a top-down analytical approach.

Another positive aspect of the results of the simulation is their contribution in obtaining the buy-in from the various partners involved in this substantial change of processing business surveys at Statistics Canada. The success of the Rolling Estimates relies on a close collaboration between the collection, processing, methodology and analytical teams.

References

Claveau, Jeannine, Joanne Leung and Claude Turmelle (2012). *Embedded Experiment for Non-Response Follow-Up Methods of Electronic Questionnaire Collection.* Proceedings of the ICES-IV Conference, Montreal, June 2012.

Godbout, Serge, Yanick Beaucage and Claude Turmelle (2011). *Quality and Efficiency using a Top-Down Approach in the Canadian's Integrated Business Statistics Program*. Conference of European Statisticians, Work Session on Statistical Data Editing. Ljubljana, Slovenia (9-11 May 2011).

Mills, Fraser, Serge Godbout, Keven Bosa, Claude Turmelle (2013), *Multivariate Selective Editing in the Integrated Business Statistics Program*, Proceedings of the Joint Statistical Meeting (JSM) Conference. Montreal, Canada.

Ravindra, Daniela (2012). *The Integrated Business Statistics Program: An Overview*. Proceedings of the ICES-IV Conference, Montreal, June 2012.

Turmelle, Claude., Serge Godbout, Keven Bosa (2012), *Methodological Challenges in the Development of Statistics Canada's New Integrated Business Statistics Program*, Proceedings of the International Conference on Establishment Surveys (ICES) IV Conference. Montreal, Canada.