The Relative Deviation from Predicted Values as a Tool to Prioritize Units for Failed Edit Follow-Up in IBSP

Matei Mireuta[1], Jessica Andrews[1] and Pierre Daoust[1]

ABSTRACT

The production of official statistics can be considerably hindered by errors in collected data. The process of correcting these errors is referred to as statistical data editing and has traditionally been an expensive and time consuming manual effort. However, it is becoming increasingly accepted in practice that correction of a small subset of influential errors is often sufficient to guarantee a good quality estimate and therefore, that additional correction of data past a certain threshold yields only minor improvements in quality. The concept of selective editing involves identifying these significant errors, estimating their impact on the final estimate and providing for a signal to halt editing operations when an error tolerance is reached. In this work, we compare several methods based on predicted values for estimating reporting errors and for prioritizing units for failed-edit follow-up in the context of the Integrated Business Statistics Program at Statistics Canada.

KEY WORDS: Active Collection, Failed Edit Follow-up, Selective Editing, Predicted Value

RÉSUMÉ

La production de statistiques officielles est souvent compliquée par la présence d'erreurs dans les données collectées. La correction de ces erreurs se fait traditionnellement par un long processus de vérification manuelle des données. Cependant, il est de plus en plus reconnu que la correction d'un petit sous-ensemble d'erreurs influentes est souvent suffisante pour garantir une bonne qualité des estimations finales et donc, que la correction au-delà d'un certain seuil n'a que peu d'impact. La vérification sélective des données s'efforce d'identifier ces erreurs influentes, d'en évaluer l'impact sur les estimations finales et de signaler l'arrêt des opérations de vérification lorsqu'un certain seuil d'erreur est atteint. Dans ce travail, nous comparerons des méthodes utilisant des valeurs prédites pour l'évaluation de l'ampleur des erreurs de réponse et pour la priorisation du suivi des unités dans le cadre du Programme Intégré de Statistiques des Entreprises de Statistique Canada.

MOTS CLÉS : Collecte active, suivi des échecs de vérification, vérification sélective, valeur prédite

1. INTRODUCTION

The Integrated Business Statistics Program (IBSP) provides a generalized model for many business surveys at Statistics Canada. This model was developed to reduce operating costs, improve quality assurance, and accelerate the delivery of new statistical programs. The IBSP system follows a non-traditional, circular process termed Rolling Estimate (Turmelle, Godbout, Bosa and Mills, 2014) whereby most survey steps are carried out in a cyclical fashion several times during a particular survey's timeline. Shortly after electronic questionnaire invitations are sent out, the first group of respondent data (along with data from administrative sources) is integrated into the Rolling Estimate model and treated as if it were the complete collection for the survey. In other words, the usual steps of editing, imputation (for total and partial non-response) and estimation are carried out and a set of preliminary estimates is obtained. Then, based on the quality of these preliminary estimates, an active collection tool prioritizes units for follow-up by collection services. Once additional data are obtained, the process is repeated and a new set of updated estimates is produced. The Rolling Estimate process is repeated in this fashion until an acceptable level of quality is achieved and the final estimates then move to the dissemination/publication

1 BSMD, 100 Tunney's Pasture Driveway, Ottawa, Ontario, K1A 0T6, matei.mireuta@canada.ca, jessica.andrews@canada.ca, pierre.daoust@canada.ca

stage. The main goal of this article is to present recent changes brought to the active collection tool (section 3) following a detailed description of its implementation (section 2).

2. ACTIVE COLLECTION TOOL

2.1 Key Estimates

The active collection tool is an algorithm that prioritizes units for collection follow-up based on the quality of preliminary estimates produced during a particular Rolling Estimate iteration. Instead of considering the entire list of estimates for a survey, the active collection tool focuses only on the quality of the most important (termed key) estimates for that survey. Therefore, only key estimates will drive collection follow-up efforts and they are chosen a priori in conjunction with subject matter experts and other stakeholders. For most business surveys in IBSP, a typical key estimate would be the total of an important variable, such as revenue, within an important province and industry class.

2.2 Factors of Importance, Quality Measures and Quality Targets

Once a list of key estimates is selected, a factor of importance (FI) is associated to each entry and serves as a measure of the importance of that particular key estimate relative to the rest of the key estimates. Additionally, a quality target is also set for each key estimate and for every quality measure used in a particular instance of the active collection tool. The quality measure is simply a tool to evaluate the quality of key estimates periodically at each Rolling Estimate iteration and the quality target is an objective or goal that the active collection algorithm will strive to achieve for each key estimate. In the current implementation, two quality measures are used simultaneously. The Key Variable Weighted Response Rate quality measure (KVWRR) is used to increase response rates, while the Relative Deviation from Predicted Values quality measure (RDPV) is used to ensure response accuracy. For KVWRR, the quality targets can be set intuitively in terms of response rate goals (or historical response rates) for each key estimate, with higher targets for more important key estimates. The methodology for setting quality targets for RDPV is, in contrast, more complex and will be presented in section 3.

2.3 Quality Indicators, Quality Distances and Measures of Impact

A quality indicator is associated to each quality measure and this simply represents the level of quality achieved at a particular time point for key survey estimates. For KVWRR, an intuitive and straightforward quality indicator function can be defined in terms of the response rate in a similar manner to the economically weighted response rate described in Godbout, Beaucage and Turmelle (2011). The focus of this manuscript (and section 3 in particular) will be to develop a quality indicator for the RDPV quality measure that strives to achieve the same level of intuitiveness. The RDPV quality measure was described in Turmelle, Godbout, Bosa and Mills (2014) where the authors suggest the use of a predicted value to quantify the accuracy of survey data. The predicted values are compared to unedited reported values, or to imputed values for nonresponse. This method can be considered as a form of selective editing. Other methods using predictions have been used previously -- c.f. Hedlin (2008) as well as de Waal (2013) for an excellent review. Selective editing methods are based on the idea that correction of all errors in the data, whether done manually or through collection follow-up, is inefficient. In practice, correction of only a small subset of influential errors has often been deemed sufficient to guarantee good quality estimates. Traditionally, a local score is assigned to each unit based on a measure of likelihood that the unit's data is erroneous combined with a measure of the influence of the unit on a relevant estimate. These scores are combined into a global score, and thresholds for stopping editing efforts are derived based on simulations using fully edited historical data (de Waal, 2013).

In Turmelle, Godbout, Bosa and Mills (2014) and consequently in this manuscript, a slightly different approach than the traditional approach described above is used. First a quality indicator will be derived and it will be interpreted as the total amount of error or total amount of deviation from predicted values within a key estimate. This indicator will then be compared to a quality target to determine if editing for a particular key estimate is deemed sufficient. A measure of impact will also be assigned to each unit in a particular key estimate and will be interpreted as the expected improvement in the quality indicator should the unit be successfully edited (i.e. its value corrected to or close to its associated predicted value). In a final step, the measures of impact will be combined into a global measure of impact which will be considered as the improvement in the quality of all key estimates should a unit be successfully edited.

As mentioned in section 1, at each Rolling Estimate iteration the data is treated as if it were the complete collection for the survey. As a consequence, all variables for all units are either reported or imputed, even for total non-response, or in other words the dataset is complete. Additionally a predicted value is available for each unit. This value exists before the start of the survey and is fixed in time unlike the imputations of non-respondents generated during a particular Rolling Estimate iteration.

The discussion and results presented here will be restricted to the level of one key estimate and thus a simpler notation than in Turmelle, Godbout, Bosa and Mills (2014) will be developed. The notion of multiple key estimates will only arise briefly at the end of this subsection (see Turmelle, Godbout, Bosa and Mills, 2014 for complete details).

For a key estimate K, let

a) S denote the sub-sample of units that contribute to K
b) y_i denote the value of the variable y pertaining to K for unit i (may be reported or imputed but not missing)
c) y_i^P denote the predicted value of the variable y allocated to K for unit i
d) w_i denote the estimation weight of unit i
e) FI_K denote the factor of importance of K

The quality indicator originally associated to the RDPV quality measure is defined in Turmelle, Godbout, Bosa and Mills (2014) as

$$QI_{RDPV}(K) = \frac{\sqrt{\sum_{i\epsilon S}(w_i(y_i - y_i^P))^2}}{\sum_{i\epsilon S} w_i y_i}. \tag{1}$$

In order to avoid the complexities of a small or negative denominator, all variables y, y^P and weights w are assumed to be positive (including 0) and at least one strictly greater than 0. Furthermore, the expected improvement in (1) obtained should a unit be successfully edited is well approximated in the case of a significant contributor to $QI_{RDPV}(K)$ by the measure of impact below

$$MI_{\mathrm{RDPV}}^K(i) = \frac{|w_i(y_i - \mathrm{y_i^P})|}{\sum_{i\epsilon S} w_i y_i}. \tag{2}$$

A quality distance (QD) is defined as a relative distance between the quality indicator calculated in (1) and a pre-set quality target. The quality distance represents the gap between the current level of quality ($QI_{\mathrm{RDPV}}(K)$) and the objective ($QT_{\mathrm{RDPV}}(K)$) that was set for K. By construction, the quality distance will be 0 if and only if the quality target is met or surpassed.

$$QD_{\mathrm{RDPV}}(K) = \begin{cases} max\left(\frac{QI_{\mathrm{RDPV}}(K) - QT_{\mathrm{RDPV}}(K)}{QI_{\mathrm{RDPV}}(K)}, 0\right) & if\ QI_{\mathrm{RDPV}}(K) > 0 \\ 0 & if\ QI_{\mathrm{RDPV}}(K) = 0 \end{cases} \tag{3}$$

A measure of impact will be obtained for unit i for each key estimate that unit i contributes to. These measures of impact scores are then combined into a global measure of impact (GMI) using the importance factors and quality distances of the key estimates the unit contributes to. As mentioned, this can be interpreted as the improvement in the quality of all key estimates if unit i were successfully edited.

$$GMI_{RDPV}(i) = \sqrt{\frac{\sum_{K\in\{Key\ Estimates\}}\left(FI_K * QD_{\mathrm{RDPV}}(K) * MI_{\mathrm{RDPV}}^K(i)\right)^2}{\sum_{K\in\{Key\ Estimates\}} {FI_K}^2}}. \tag{4}$$

Ultimately, the goal of the active collection tool is to efficiently prioritize units for collection follow-up through the use of a single priority list. The KVWRR quality measure is used in the active collection tool to drive non-response follow-up (NRFU) efforts while the RDPV quality measure is used to drive failed-edit follow-up (FEFU) efforts. Each unit will have one GMI for each of the two quality measures, which will then be balanced to determine the unit's position on the priority list.

3. RELATIVE DEVIATION FROM PREDICTED VALUES QUALITY MEASURE

This section will describe the development of the RDPV quality measure that is currently in use in IBSP. The RDPV quality indicator described in (1) presents several challenges that will become apparent in the remainder of this section. Our goal was to build upon the framework presented in section 2.3 by exploring additional quality indicators in an attempt to overcome the disadvantages of (1). The approach presented in this section will be very similar to the one described in Turmelle, Godbout, Bosa and Mills (2014), but another formula for the RDPV quality indicator will ultimately be suggested.

3.1 RDPV Quality Indicator

Our main objective was to develop a quality indicator with the following desirable properties:

a) The quality indicator should be related to a notion of error for a key estimate K
b) The quality indicator should admit a straightforward interpretation and a straightforward setting of quality targets
c) The quality indicator should be a decreasing function following successful editing, at least when y_i^P is exactly accurate, i.e. when y_i^P is equal to the true value of variable y_i.

The following three indicator functions were explored:

$$QI_0 = \frac{\sqrt{\sum_{i\epsilon S}(w_i(y_i - y_i^P))^2}}{\sum_{i\epsilon S} w_i y_i}, \qquad QI_1 = \frac{\sum_{i\epsilon S} |w_i\left(y_i - y_i^P\right)|}{\sum_{i\epsilon S} w_i y_i}, \qquad QI_2 = \frac{\left|\sum_{i\epsilon S} w_i\left(y_i - y_i^P\right)\right|}{\sum_{i\epsilon S} w_i y_i}.$$

The indicator QI_0 was presented in (1) and has the advantage of placing more emphasis on larger (or exceptional) deviations. QI_1 and QI_2 are alternatives that strive to mitigate the challenges inherent to QI_0.

3.1.1 Absolute Total Error of Response/Imputation

At a given Rolling Estimate iteration, the error in a key estimate total will be due to both errors of response (for respondents) and errors of imputation (for non-respondents). In other words, we will assume each value y_i can be written as the sum of the true value y_i^T of unit *i* and an error of response/imputation e_i. Then the key estimate total can be written as

$$\sum_{i\epsilon S} w_i y_i = \sum_{i\epsilon S} w_i(y_i^T + e_i) = \sum_{i\epsilon S} w_i y_i^T + \sum_{i\epsilon S} w_i e_i. \tag{5}$$

For the moment, we take a broader view on editing that does not assume a specific method for correcting the error term in (5). We will consider until section 3.2 that a successful editing operation could be achieved through any method, not just via collection follow-up of respondent units. Therefore, we are interested in (5) in the error from both sources, that is the error in the key estimate total that is obtained from using erroneous unit values as opposed to the true unit values.

A bound for the second term of (5) will yield a bound for the key estimate total calculated based on the true values y_i^T. Indeed, if a value c can be found such that the absolute total error of response/imputation is less than c at a particular Rolling Estimate iteration, i.e.

$$\frac{|\sum_{i\epsilon S} w_i e_i|}{\sum_{i\epsilon S} w_i y_i} \leq c \tag{6}$$

then the following holds

$$(1+c)\sum_{i\epsilon S} w_i y_i \geq \sum_{i\epsilon S} w_i y_i^T \geq (1-c)\sum_{i\epsilon S} w_i y_i. \tag{7}$$

The value c in (7) is the maximum change or improvement in the current key estimate total ($\sum_{i\epsilon S} w_i y_i$) that would be obtained if all errors were corrected. Since multiple methods of data correction can occur simultaneously, c is also the maximum change in the key estimate total that can be achieved by either of these methods, including successful follow-up of all respondent units having failed edits.

If y_i^P is expressed in terms of the same value y_i^T and an error of prediction e_i^P then, using the triangle inequality, the following is true for QI_1 and QI_2

$$QI_1 = \frac{\sum_{i\epsilon S} |w_i\left(y_i - y_i^P\right)|}{\sum_{i\epsilon S} w_i y_i} \geq QI_2 = \frac{\left|\sum_{i\epsilon S} w_i\left(y_i - y_i^P\right)\right|}{\sum_{i\epsilon S} w_i y_i} = \frac{\left|\sum_{i\epsilon S} w_i\left(y_i^T + e_i - y_i^T - e_i^P\right)\right|}{\sum_{i\epsilon S} w_i y_i} = \frac{\left|\sum_{i\epsilon S} w_i\, e_i - \sum_{i\epsilon S} w_i\, e_i^P\right|}{\sum_{i\epsilon S} w_i y_i} \quad (8)$$

Furthermore, if the sum of response/imputation errors is assumed much larger than the sum of prediction errors, then the second term in the numerator of (8) can be ignored and QI_1 becomes an estimated bound while QI_2 becomes an estimate for (6). The interpretation of both indicators would be identical to the value c in (7), i.e. the maximum change in the key estimate total that can be achieved if all units are successfully followed-up and corrected. This interpretation leads to straightforward quality targets. For example, it may be desired to follow up units until a point when further follow-up would only be expected to change the key estimate total by a maximum of say 5%.

Conversely, the quality indicator QI_0 is not a bound for the absolute total error (6) and lacks a straightforward interpretation under this framework.

The assumption that the sum of prediction errors is negligible may initially appear strong. It can be argued that this assumption may be reasonable when the prediction model is unbiased. It is also noteworthy to mention that QI_1 is a more conservative bound for (6) than QI_2, thereby providing some protection against deviations from this hypothesis. A second interpretation of both indicators QI_1 and QI_2 is possible that does not make explicit use of the above assumption.

3.1.2 An Average Deviation from Prediction

The indicator QI_1 is a sum of all deviations from prediction, where some deviations are large and others are small. An interesting interpretation arises when this sum is re-distributed to each y_i as a constant proportion of y_i. In other words, for any QI_1, a value of α is sought such that

$$QI_1 = \frac{\sum_{i\epsilon S} |w_i\left(y_i - y_i^P\right)|}{\sum_{i\epsilon S} w_i y_i} = \frac{\sum_{i\epsilon S} |w_i\left(y_i - \alpha y_i\right)|}{\sum_{i\epsilon S} w_i y_i}. \quad (9)$$

As mentioned in section 2.3, $w_i \geq 0$ and $y_i \geq 0$. Solving for α in (9) is then straightforward yielding

$$\alpha = 1 \pm QI_1. \quad (10)$$

The two equations above imply that the quality measured by QI_1 is equivalent to the quality obtained had all predicted values been either QI_1 above or QI_1 below each y_i. Under this interpretation, QI_1 may be thought of as an "average" deviation from prediction and quality targets can in turn be set accordingly. For example, it may be desired to follow up units until a point when this average deviation from prediction is below a tolerated level of say 10% of y_i.

The point of view taken in this section is different from the one in section 3.1.1. The sum of prediction errors are not necessarily assumed negligible, but instead the predicted values represent a benchmark for comparison and quality targets may be set historically. For instance, if a key estimate total, after extensive follow-up and manual editing, was published within 10% of prediction in a previous cycle, then a quality target of 10% is a reasonable tolerance if conditions are expected to remain the same in the current cycle.

A second version of (10) can easily be obtained if the values y_i are instead written as a proportion of y_i^P and this yields the same interpretation, only with y_i^P as the reference point. Additionally, this reasoning will yield a similar interpretation in the case of QI_2. On the other hand, when this principle is applied to QI_0, the result becomes more difficult to interpret

$$QI_0 = \frac{\sqrt{\sum_{i\epsilon S}(w_i(y_i - y_i^P))^2}}{\sum_{i\epsilon S} w_i y_i} = \frac{\sqrt{\sum_{i\epsilon S}(1-\alpha)^2 (w_i y_i)^2}}{\sum_{i\epsilon S} w_i y_i} = \frac{|1-\alpha|\sqrt{n * (Var(wy) + \overline{wy}^2)}}{n * \overline{wy}} = |1-\alpha|\sqrt{\frac{CV(wy)^2 + 1}{n}} \quad (11)$$

where $\overline{wy} = \frac{\sum_{i\epsilon S} w_i y_i}{n}$, $Var(wy), CV(wy)$ represent respectively the average, variance and coefficient of variation for the sample of points $w_i y_i$.

The result in (11) shows that QI_0 does not only capture an "average" deviation α, but also the dispersion of the products $w_i y_i$ as well as the number of units n contributing to *K*. This added complexity renders setting quality targets more difficult as particular care must be used to ensure that the dispersion of units and the number of units is adequately accounted for a given tolerated deviation α.

3.1.3 Monotonicity of Quality Indicators

In addition to a straightforward interpretation, the quality indicator for RDPV should ideally be a monotone function with respect to successful editing operations. Since the RDPV quality indicator is a measure of the total deviation from prediction, the correction of a reported value should lead to a decrease in the quality indicator, at least under the assumption that the predicted value equals the true value for each unit.

For a unit *m*, it can be shown by example that QI_0 is not monotone decreasing even when $y_m^P = y_m^T$ and the reason follows from (11). Indeed, a successful editing operation may reduce the average deviation α but may at the same time increase the dispersion of values as measured by $CV(wy)$. It is also straightforward to show that QI_2 is not monotone decreasing.

On the other hand, QI_1 is monotone decreasing under certain conditions. If $y_m^P = y_m^T$, then a successful correction would update y_m by the true value y_m^P (equal y_m^T) and the following would need to hold.

$$QI_1 \geq QI_1(after\ editing\ y_m)$$

$$\frac{\sum_{i\epsilon S}|w_i\,(y_i - y_i^P)|}{\sum_{i\epsilon S} w_i y_i} \geq \frac{\{\sum_{i\epsilon S}|w_i\,(y_i - y_i^P)|\} - |w_m(y_m - y_m^P)|}{\{\sum_{i\epsilon S} w_i y_i\} - w_m y_m + w_m y_m^P}$$

$$\left\{\sum\nolimits_{i\epsilon S}|w_i(y_i - y_i^P)|\right\}\left(w_m(y_m - y_m^P)\right) \leq \left\{\sum\nolimits_{i\epsilon S} w_i y_i\right\}|w_m(y_m - y_m^P)|$$

$$QI_1 \geq \frac{|w_m(y_m - y_m^P)|}{\left(w_m(y_m - y_m^P)\right)} = -1\ \ (if\ y_m < y_m^P)\ \ or\ \ QI_1 \leq \frac{|w_m(y_m - y_m^P)|}{\left(w_m(y_m - y_m^P)\right)} = 1\ (if\ y_m > y_m^P). \quad (12)$$

Given that $w_m,\ y_m$ and y_m^P are all positive, then $QI_1 \geq 0$ by construction and therefore the first bound in (12) is always satisfied. However, in order to maintain monotonicity under both conditions QI_1 needs to be bounded above by 1.

When $y_m^P \neq y_m^T$, we have found that monotonicity is still preserved in most practical cases although specific examples can be created where is does not hold.

The indicator QI_1 satisfies the three objectives set out in section 3.1 and therefore was retained as the quality indicator for RDPV.

$$QI_{RDPV}(K) = min\left(\frac{\sum_{i\epsilon S}|w_i\,(y_i - y_i^P)|}{\sum_{i\epsilon S} w_i y_i}, 1\right). \quad (13)$$

3.2 RDPV Measure of Impact

The measure of impact for RDPV will be defined in a similar manner as in section 2.3. It will be the difference induced in the quality indicator following a successful editing operation, under the assumption $y_m^P = y_m^T$ and $QI_1 < 1$.

$$QI_1 - QI_1(after\ editing\ y_m) = \frac{\{\sum_{i\epsilon S, i\neq m} |w_i\left(y_i - y_i^P\right)|\} + |w_m(y_m - y_m^P)|}{\sum_{i\epsilon S, i\neq m} w_i y_i + w_m y_m} - \frac{\{\sum_{i\epsilon S, i\neq m} |w_i\left(y_i - y_i^P\right)|\}}{\sum_{i\epsilon S, i\neq m} w_i y_i + w_m y_m^P}$$

$$= \frac{\{\sum_{i\epsilon S, i\neq m} |w_i\left(y_i - y_i^P\right)|\}\left(w_m(y_m^P - y_m)\right) + \{\sum_{i\epsilon S, i\neq m} w_i y_i + w_m y_m^P\}|w_m(y_m - y_m^P)|}{\{\sum_{i\epsilon S, i\neq m} w_i y_i + w_m y_m\}\{\sum_{i\epsilon S, i\neq m} w_i y_i + w_m y_m^P\}}$$

$$= \frac{|w_m(y_m - y_m^P)|}{\{\sum_{i\epsilon S, i\neq m} w_i y_i + w_m y_m\}} + \frac{\{\sum_{i\epsilon S, i\neq m} |w_i\left(y_i - y_i^P\right)|\}\left(w_m(y_m^P - y_m)\right)}{\{\sum_{i\epsilon S, i\neq m} w_i y_i + w_m y_m\}\{\sum_{i\epsilon S, i\neq m} w_i y_i + w_m y_m^P\}}. \quad (14)$$

The second term in (14) is in general smaller than the first and will be ignored for computational purposes. This will then yield the measure of impact described in Turmelle, Godbout, Bosa and Mills (2014) with an addition of an upper bound of 1 so as to restrict its range to that of QI_{KVWRR}.

$$MI_{RDPV}^{K}(i) = min\left(\frac{|w_i(y_i - y_i^P)|}{\sum_{i\epsilon S} w_i y_i}, 1\right). \quad (15)$$

As a side note, in addition to the approximate change induced in QI_{RDPV}, it can easily be shown that MI_{RDPV}^{K} is also the anticipated absolute change in the key estimate total $\sum_{i\epsilon S} w_i y_i$ should an editing operation on unit *m* be successful and should $y_m^P = y_m^T$.

$$\frac{\left|\{\sum_{i\epsilon S, i\neq m} w_i y_i + w_m y_m\} - \{\sum_{i\epsilon S, i\neq m} w_i y_i + w_m y_m^P\}\right|}{\sum_{i\epsilon S} w_i y_i} = \frac{|w_m(y_m - y_m^P)|}{\sum_{i\epsilon S} w_i y_i}. \quad (16)$$

It is important to mention that equations (12), (14) and (16) are expected to hold as long as $y_m^P \approx y_m^T$ and will inevitably fail if predictions are significantly different from the corresponding true values.

The quality measured given by equation (13) and the impact measured by equation (15) do not assume any underlying method of editing at this point. In IBSP, the successful correction of a unit's value may be achieved through several means that often occur simultaneously. For example, a unit's imputed value may be corrected by the Rolling Estimate process, a unit's reported value may be manually corrected or a unit's imputed value may even be rectified through a correct reported value. In the current implementation of the active collection tool, the quality indicator in equation (13) is calculated using the whole sample and therefore, failed edit follow-up is merely one of the means by which the quality indicator may be decreased. This allows situations where, for instance, extensive manual editing may bring a key estimate below its quality target and eliminate the need for failed-edit follow-up (FEFU) altogether or vice versa. On the other hand, the measure of impact in equation (15) will be restricted to respondent units that are eligible for collection follow-up, since the interest here is only in the impact on equation (13) of units that can be corrected through successful FEFU.

Finally, the quality distance and global measure of impact for RDPV will follow equation (3) and equation (4) as in section 2.

4. CONCLUSION

Analysis of several IBSP surveys revealed that monotonicity problems of QI_0 occurred frequently and simulations using realistic data supported the change to QI_1. Therefore, the Relative Deviation from Predicted Values quality measure as described in section 3 was implemented in 2017 in conjunction with the Key Variable Weighted Response Rate quality measure mentioned in section 2. In this implementation, the predicted values were obtained mainly via historical imputation or via ratio or mean imputation for units where no historical data existed. Given the limited studies currently available on the potential bias of this prediction model, the interpretation of the quality indicator defined in section 3.1.2 was favored over the interpretation of section 3.1.1. The quality targets for the present cycle were set based on deviations from prediction achieved on final estimates of the previous cycle.

ACKNOWLEGEMENTS

We would like to thank Rubina Singla for her work on exploring quality indicator functions for several IBSP surveys, as well as Keven Bosa, Serge Godbout and Elisabeth Neusy for reviewing this paper and for their helpful comments.

REFERENCES

Godbout, S., Beaucage, Y. and Turmelle, C. (2011). Achieving Quality and Efficiency Using a Top-Down Approach in the Canadian Integrated Business Statistics Program. *Conference of European Statisticians*. Ljubljana, Slovenia.

Hedlin, D. (2008). Local and Global Score Functions in Selective Editing. *UN/UNECE*. Vienna.

Turmelle, C., Godbout, S., Bosa, K. and Mills, F. (2014). A Quality Driven Approach to Managing Collection and Analysis. *Proceedings of the European Conference on Quality in official statistics*. Vienna.

de Waal, T. (2013). Selective Editing: A Quest for Efficiency and Data Quality. *Journal of Official Statistics*, 473–488.