UNIVERSITY OF
Southampton
Southampton Statistical
Sciences Research Institute

Working Paper M08/10

Methodology

A Generic Tool To Assess Impact Of Changing Edit Rules In A Business Survey-An Application To The UK Annual Business Inquiry Part 2

Pedro Luis Do Nascimento Silva, Rober Bucknall, Ping Zong, Alaa Al-Hamad

Abstract

Business surveys often use complex sets of edit rules (edits, for short) to check returned questionnaires (records), locate suspicious or unacceptable responses, and support data cleaning operations prior to using the survey responses for estimation of the required target parameters. These sets of edits are complex because they may involve large numbers of survey questionnaires and variables, they may contain a large number of edits, and the edits may depend on a large number of tolerance parameters. When such sets of edits are used, they may cause large numbers of record failures and generate substantial costs of revision, especially if edit failures are dealt with by means of clerical operations, like reviewing original paper questionnaires or digital images of these, and re- contacting businesses for clarification and/or correction of the responses provided. Costs can be high both in terms of the resources required, as well as in terms of timeliness of survey processing, by delaying availability of the survey data for estimation and publication.

In this paper we describe a generic tool, developed as a result of the collaboration between the University of Southampton and the ONS. This tool can help to assess the potential impact of changing the edits in a specified business survey. It is a SAS macro using the IML language which enables calculation of a number of edit performance and data quality indicators. Changes to the set of edits aiming to ‘relax’ the existing edits so that failure rates decrease and efficiency savings are achieved are assessed by means of several edit- related performance indicators, like failure and hit rates, false hit rates, etc.. Data quality indicators include proportion of errors missed and estimates of the bias resulting from missing errors for a specified revision of the set of edits. Edit designers and managers can then aim to fine tune their edits so that failure rates, false hit rates and editing costs are reduced, while data quality is preserved. An illustration is provided by the application of the tool to revise the edits used for the UK Annual Business Inquiry Part 2 to the reference year 2007.

A GENERIC TOOL TO ASSESS IMPACT OF CHANGING EDIT RULES IN A BUSINESS SURVEY – AN APPLICATION TO THE UK ANNUAL BUSINESS INQUIRY PART 2

Pedro Luis do Nascimento Silva, University of Southampton, United Kingdom
Robert Bucknall, Ping Zong and Alaa Al-Hamad, Office for National Statistics, United Kingdom

Abstract

Business surveys often use complex sets of edit rules (edits, for short) to check returned questionnaires (records), locate suspicious or unacceptable responses, and support data cleaning operations prior to using the survey responses for estimation of the required target parameters. These sets of edits are complex because they may involve large numbers of survey questionnaires and variables, they may contain a large number of edits, and the edits may depend on a large number of tolerance parameters. When such sets of edits are used, they may cause large numbers of record failures and generate substantial costs of revision, especially if edit failures are dealt with by means of clerical operations, like reviewing original paper questionnaires or digital images of these, and re-contacting businesses for clarification and/or correction of the responses provided. Costs can be high both in terms of the resources required, as well as in terms of timeliness of survey processing, by delaying availability of the survey data for estimation and publication.

In this paper we describe a generic tool, developed as a result of the collaboration between the University of Southampton and the ONS. This tool can help to assess the potential impact of changing the edits in a specified business survey. It is a SAS macro using the IML language which enables calculation of a number of edit performance and data quality indicators. Changes to the set of edits aiming to 'relax' the existing edits so that failure rates decrease and efficiency savings are achieved are assessed by means of several edit-related performance indicators, like failure and hit rates, false hit rates, etc.. Data quality indicators include proportion of errors missed and estimates of the bias resulting from missing errors for a specified revision of the set of edits. Edit designers and managers can then aim to fine tune their edits so that failure rates, false hit rates and editing costs are reduced, while data quality is preserved. An illustration is provided by the application of the tool to revise the edits used for the UK Annual Business Inquiry Part 2 to the reference year 2007.

Keywords: editing; survey quality; survey process; generic software; quality assessment.

1. Introduction And Background

Business surveys often use a set of edit rules to check returned questionnaires (records), locate suspicious or unacceptable responses, and support data cleaning operations prior to using the survey responses for estimation of the required target parameters. It is not uncommon for some of these sets of edit rules (edits for simplicity) to become complex because:

a) They may involve large numbers of survey variables (questions or record fields);
b) They may contain a large number of edits (tests or validation rules in the ONS jargon);
c) The edit rules may depend on a large number of tolerance parameters (gates in the ONS jargon);
d) The edit rules may be split into two subsets – one comprising so-called fatal or hard edits, i.e. edits that must be satisfied by all records; and one comprising query or soft edits, namely edits that flag records for verification, which are useful in locating suspicious or surprising observations, but which may or may not indicate an error.

It is also not uncommon to find sets of edits which are in use, but which were not specifically designed for the current survey period, having resulted from historical or legacy systems which were put in place long ago, and evolved with little or no change for a number of survey periods (often years). When such sets of edits are used, they may cause large numbers of failures and generate substantial costs of revision, especially if edit failures are dealt with by means of clerical operations like reviewing original paper questionnaires or digital images of these, and/or re-contacting businesses for clarification and/or correction of the responses provided. Costs can be high both in terms of the resources required, as well as in terms of timeliness of survey processing, by delaying availability of the survey data for estimation and publication.

In an attempt to achieve efficiency savings, survey managers may try and 'relax' the existing edits so that failure rates decrease. It would be ideal if the revised edits resulted in smaller failure rates, but without any noticeable impact on the quality of the resulting survey data. Three approaches for survey edit revision in use at ONS are:

a) Filtering or sub-setting – this comprises introducing a record filter which selects the records to be submitted to the full set of edits; records not selected by the filter are not submitted to validation; often the filter condition is related to the size of the unit corresponding to the record; ONS uses the so-called edit 'pre-conditions' to modify the edits in order to implement record filtering;
b) Gate widening – this consists in revising the tolerance parameters (gates) in individual edit rules, such that flagging of suspicious records for revision is less frequent than with previously used values;
c) Edit deletion – this consists of simply discarding some of the edits previously used to flag suspicious records.

A common feature of all three approaches is the fact that, using the revised set of edits, some records which would have been flagged under the previously used set of edits are no longer flagged, hence are not submitted to validation (or clerical verification), and their corresponding *raw data* should be used for survey estimation and publication. If it is true that this can lead to cost-savings during the costly validation operation, the corresponding impact on survey data quality is not so obvious. To enable informed decision making in similar scenarios, the ONS developed the "decision support tool" SNOWDON Al-Hamad, Martín and Brown (2006), which aims to "… *help survey managers evaluate what savings can be achieved, at what cost to output quality, across many alternative permutations of editing rule parameters.*"

This tool was initially applied to a relatively small quarterly survey, the Quarterly Stocks Inquiry. This survey is small in the sense that the number of survey variables and targets for inference is small, and consequently, the corresponding set of edits is also small. When the SNOWDON tool was considered for application to the Annual Business Inquiry Part 2 (ABI2), it became clear that its methodology and corresponding software implementation required substantial extension to enable handling of the much more complex survey scenario, with hundreds of variables, in different questionnaire types, and a much bigger set of edits. This extension and the resulting revised SNOWDON tool, which we labelled SNOWDON-X are described in this paper.

Before we move on to discuss this tool, first let us examine these edit revision approaches more closely. The first approach, *filtering*, has been recently used in the Annual Business Inquiry – Part 2 (ABI2) to achieve savings in the validation operation. In 2005 a filter was introduced that selected for validation only records corresponding to businesses with a returned turnover of 250 thousand pounds and above. This filter was implemented as a 'pre-condition' for all edits affecting specified questionnaire types (form types). For example, in the catering sector, there are 18 edits which apply to the short questionnaire (form=CT205). A pre-condition of (Q346 > 249, i.e. the returned turnover Q346 in thousand pounds is greater than 249) was added to each of these edits. This way of filtering is more flexible than a global record filter, because one could decide to filter records for some edits but not for others. However, if it is to be applied globally (i.e. to all edits), it would be more efficient to use a simple record filter.

A detailed assessment of the overall impact of introduction of this filtering procedure is not available. However, for the short questionnaires in the catering sector, there was a 45.7% reduction in the number of records flagged for validation, and a 52.7% reduction in the number of edit failures to be verified by the validation staff – see Silva (2007). These reductions offer only a crude indication of the potential savings, since records excluded from the standard application of micro-editing rules, and hence left unchanged at this stage, may later be flagged as suspicious or erroneous during the output editing (macro-editing) operation performed prior to estimation and publication. Unfortunately, no information is available to assess the impact of the filtering on the macro-editing (or output editing) of the ABI2 in

2005. But it has been demonstrated that if no further action was taken on the records for small businesses the negative impact on the quality of the estimates could be substantial – see Bucknall (2006).

The second approach – gate widening – is considered as the next step in search of additional savings for the costly validation operation of the ABI2 survey. To implement this initiative ONS considered using the SNOWDON tool Al-Hamad et al. (2006) to assist in defining revised tolerance parameters for the ABI2 edits. However the methodology and software needed to be extended substantially in order to deal effectively with the ABI2 survey. The main reasons why such an extension was required are: enabling the tool to deal with multivariate surveys, improve the bias estimation procedures, and improve the analytic potential of the tool by computing a number of additional measures of savings and impact of the changes to the edit rules.

In this context, we briefly describe the SNOWDON approach in section 2, discuss its limitations to deal with the ABI2 in section 3, and present the key ideas behind its extension and improvement to deal with the perceived limitations in section 4. Section 5 provides an example of how the extended approach was used in the context of ABI2 for reviewing the edits. Section 6 contains some conclusions and a discussion of the approach.

2. The Snowdon Approach And Some Terminology

SNOWDON is a name given to a SAS macro developed within ONS Methodology Directorate to estimate efficiency and quality indicators for a given survey where the current set of edits is to be revised by filtering, edit deletion and/or gate widening. The SNOWDON approach is described in detail by Al-Hamad et al. (2006). It works by assuming that the revised set of edits with the new tolerance parameters is used for validation instead of the current set of edits, in a context where previously unedited and edited survey data are available. It then uses these data to estimate a set of efficiency and quality indicators which includes:

a) the expected savings in validation;
b) the expected number of errors missed; and
c) the expected impact on quality.

The basic idea is as follows. Consider a survey which aims to estimate totals and related parameters for a specified set of domains. Let $U = \{1, 2, \ldots, N\}$ denote the set of units in the target population, uniquely labelled by the integers 1 to *N*. The survey selects a probability sample $s \subset U$ of size *n*, and observes a subset $r \subset s$ of size *m* from the target population *U*. Data for all units in *r* are available for survey processing. In order to estimate the quantities of interest, SNOWDON requires two versions of the survey data: one copy of the survey records at 'point of capture', namely before any validation took place (*raw data*); and a copy of the survey records at 'point of estimation', namely, after all required validation took place and the records are ready to be used for survey estimation and publication (*edited data*). A crucial underlying assumption is that the edited data are 'clean' or 'error free', having been submitted to the best possible validation efforts in a recent edition of the survey.

These two datasets are then used as follows. First, the current set of edits is applied to the *raw data*, and the following quantities are computed:

A = number of records failing at least one edit, i.e. *record failure count*;

B = number of records which failed at least one edit <u>and</u> were modified during validation (obtained by comparing the failed business records in the raw and edited datasets), i.e. *record hit count*.

The ratio of *A* to *m* is called the *current failure rate*, and the ratio of *B* to *A* is called the *current hit rate*.

Then the edits are modified by relaxing some of the tolerance parameters (gates), and the revised set of edits is applied to the raw data. The quantities A^* and B^* are computed, namely the revised record failure and hit counts, as well as the corresponding failure and hit rates obtained using the revised set of edits. The difference

$$\text{Savings} = A - A^* \qquad (1)$$

provides a gross measure of savings due to the edit revision, namely the number of records which would not be edited if the edit rules are revised as indicated. This measure can be translated into monetary value if an estimate of the average validation cost per record is available.

Corresponding to this level of savings, there is a corresponding impact on the quality of survey estimates measured by the following two indicators:
Number of missed errors = $B - B^*$ (2)

$$\text{RelBias} = \left| \sum_{i \in r \cap f} w_i \left(z_i - y_i \right) \right| \Bigg/ \sum_{i \in r} w_i y_i \qquad (3)$$

where $r \cap f$ is the subset of records that fail edits before they are changed but pass after edit revision, w_i is the survey weight of unit *i*, z_i is the raw response for unit *i* and y_i is the corresponding edited value. If a record does not fail any edits then the edited value is assumed equal to the initial (raw) value.

Equation (3) is a crude estimate of the absolute value of the relative bias due to errors missed relative to the weighted estimates obtained under full validation. Note that all respondent records contribute to the denominator, whereas only records with missed errors, i.e. those records that fail current edits and are changed but pass all edits after edit revision (hence would not be changed under the revised set of edits) contribute to the numerator of the absolute relative bias expression (3). All these indicators can be computed for relevant domains of analysis and/or publication, as described in Al-Hamad et al. (2006).

3. Some Limitations Of Snowdon To Deal With The ABI2

The first major limitation is the fact that the main quality measure for the relative bias is univariate, namely, it refers to a single variable. This was not a problem when SNOWDON was first implemented in the Quarterly Stocks Inquiry, because this survey has a single target variable, namely stocks. However, ABI2 is quite different, because it collects information on a large number of variables, and hence, a univariate measure of relative bias is inadequate.

The second main limitation is the form of the relative bias estimator proposed by Al-Hamad et al. (2006). This estimator can be volatile, because if two large missed errors cancelled out, the relative bias estimated using (3) would be zero, although the errors themselves could be large, and could produce substantial bias on domain estimates for which the corresponding erroneous records did not contribute together or as a pair.

Another limitation arises from the fact that SNOWDON only computes gross measures of savings, in terms of numbers of records which would not need to undergo micro-editing (or validation). Although this is an important measure of overall potential savings, this measure alone cannot explain where the savings are coming from. Suppose that the tolerance parameters are changed in two edit rules. How much of the savings are coming from the change in detection power of each edit rule? This type of information would be essential to drive the necessary survey process improvements required to make real savings in the future, namely addressing quality issues in data collection and capture which may prevent obtaining suspicious or erroneous answers in the first place. Also, if the edit sets are quite large, it may be quite challenging to set values for the tolerance parameters in several edit rules at the same time and to predict the potential impact of the changes. Hence we advocate that additional measures of edit performance are needed to make the process work.

4. Extending Snowdon To Deal With The Perceived Limitations

In order to describe the proposed extensions to SNOWDON required to deal with the perceived limitations discussed in section 3, some additional notation is required. A framework similar to that proposed in Appendix C of Luzi et al. (2007) is adopted here. Chambers (2001) provides additional guidance on the various measures used to assess performance of the set of edits under analysis.

Let #(A) denote the cardinality of the set A. Then we have $N = \#(U)$, $n = \#(s)$ and $m = \#(r)$. Let $I(\bullet)$ denote the indicator function taking the value 1 if its argument is true and zero otherwise. Denote by z_{ij} the value of the initial (raw, unedited) response recorded for variable j in record i, and by y_{ij} the value of the final (edited) response recorded for variable j in record i, j=1,…,p. Define:

$E = \{E_1, E_2, \ldots, E_L\}$ as the set of all current edits;

$F \subset E$ as the set of all current fatal (or hard) edits, which must be satisfied by all records; and

$Q \subset E$ as the set of all current query (or soft) edits, used to flag records for validation, and which may or may not indicate an error.

It follows from the above definitions that $F \cap Q = \varnothing$, $E = F \cup Q$. Let $L = \#(E)$ and $K = \#(F)$. Then it also follows that $\#(Q) = L - K$. Note that for some applications either F or Q may be empty sets, but both cannot be simultaneously empty, i.e. we assume that $E = F \cup Q \neq \varnothing$.

In addition, define the following series of detailed indicator variables:

$e_{ik} = I(\text{record } i \text{ fails edit } E_k)$;

$c_{kj} = I(\text{variable } j \text{ is active or involved in edit } E_k)$;

$r_{ij} = I(\text{variable } j \text{ is available (not missing) for record } i)$;

$v_{ij} = I\begin{pmatrix}\text{variable } j \text{ in record } i \text{ is flagged as erroneous} \\ \text{or suspicious by at least one edit in } E\end{pmatrix}$;

$g_{ij} = I(z_{ij} \neq y_{ij})$, i.e., the indicator that variable j in record i was changed during validation.

It follows from the above definitions that:

$$v_{ij} = 1 - r_{ij} \prod_{k \in E} (1 - e_{ik} c_{kj}).$$

It also follows that some relevant record level indicators can be defined using the various field level indicators defined above:

$v_i = 1 - \prod_{j=1}^{p} (1 - v_{ij})$ is the indicator that at least one field in record i is flagged as erroneous or suspicious by at least one edit in $E = F \cup Q$;

$g_i = 1 - \prod_{j=1}^{p} (1 - g_{ij})$ is the indicator that at least one variable in record i changed during editing;

$r_i = \prod_{j=1}^{p} r_{ij}$ is the indicator that record i is complete (has no missing values);

$\tilde{v}_i = 1 - v_i$ is the indicator that the record is 'clean' under the edits in E;

$\tilde{g}_i = 1 - g_i$ is the indicator that the record is unchanged after the editing;

$\tilde{r}_i = 1 - r_i$ is the indicator that the record is incomplete, i.e. has at least one missing variable.

In all the above notation, a star superscript is added to indicate that the edits considered to obtain the corresponding outcome are the revised edits E^* obtained by relaxing at least one edit in the set E. Note that edits can only be relaxed by filtering, edit deletion, gate widening or modifications which make the edit set less stringent.

The limitations related to the use of a single measure of savings in SNOWDON can be addressed by defining a number of additional indicators of the performance of a set of edits. Table 1 in the Appendix defines some general descriptive summaries based on the record level indicators defined above, which are unaffected by the changes in the current edit set. They are proposed to provide a summary of the impact of editing to the available data set in a prior edition of the survey, and to enable the analyst to have an idea of how much editing would take place in the survey if no changes are made to the edits.

Table 2 in the Appendix presents a cross-classification of data value status and error flag for variable *j* in record *i.* The idea for this table is to make clear the conceptual basis supporting definition of various edit quality and efficiency indicators, which follows. Despite recognising that this is a strong assumption, data available from a previous survey round which went through validation using the current set of rules (*E*) are considered 'clean' or error-free at the end of the validation operation for that survey. Hence raw data which were changed during validation are considered erroneous. Data can either be flagged or not. Flagging erroneous data and not flagging unchanged data are the correct decisions, and flagging data which are not in error, or not flagging data which are found to be wrong are incorrect decisions. Any editing procedure improvement should aim to increase the proportion of times that correct decisions are taken, namely detecting erroneous records, and letting undetected only those records which contain no errors.

Table 3 presents the definitions for several quality and efficiency indicators relating to individual edits, records or variables, which may be used for detailed analysis and to assess edit performance, by identifying different types of issues which may be associated with specific variables. Note that the intended use of the various indicators defined here is to help with the choice of which edit rules to target for revision and to guide the definition of alternative tolerance parameters. However, many of these indicators can also be used to assess performance of the initial edit rules in their own right. These indicators allow one to check what the problem may be with a variable: it may contain many errors which are missed by the edits, or alternatively, it may be flagged for some records where there are no errors, and thus may help identify cases of edit inefficiency.

In addition, Table 4 contains the extended set of indicators proposed to assess the impact of changes to the set of edit rules. Note that only changes in the direction of increasing tolerance of edit rules can be assessed using the approach proposed here. Given the context and to preserve some simplicity, weighted versions of various indicators were not considered. Weights were only taken into account when computing the estimates of bias due to errors missed as a result of changing the edits, as will be seen later. However, if required, it would not be difficult to extend the definitions of various other descriptive indicators, where appropriate, to incorporate record or case weights.

The quality measures proposed to extend those considered under the initial SNOWDON approach include two different measures of bias. The first one, which we call '*local bias*' and estimate for each variable using (C4) – see Table 4 in the Appendix, is the 'natural' extension of (3) modified to avoid the issue raised by the possibility of cancelling errors. However, (C4) is still a univariate measure, so we consider two alternative summaries given by (C6) and (C7). These two summaries provide complementary indicators, because (C6) is an average, and hence, may be influenced by the total number of variables involved in the edits, whereas (C7) is a maximum over the set of variables involved, and hence provides information on the 'worst case scenario' for any individual variable involved in the edits.

Note that the main difference between (C4) and (3) is that (C4) will not be affected if errors missed cancel out as may happen with (3). In addition, for any single variable, it may be shown that (3) $\leq$ (C4). Hence if the relative absolute local bias (C4) for a variable *j* is small, it follows that the absolute relative bias will also be small for the estimates of total based on the records containing the missed errors for this variable. In this sense, (C4) is a more conservative indicator than (3).

The other difference between our extended set of indicators and (3) is that (C6) is an average relative bias over the set of *p* variables considered in the survey edits, hence it is multivariate by definition, and accounts for bias on all variables involved in the edits. If we wanted to be conservative, we could use instead the maximum relative local bias (C7) to summarize the local bias for the variables involved in the editing.

These indicators of bias did not provide a sufficiently comprehensive picture. Hence the additional concepts relating to a 'global' measure of bias were introduced. The reason behind this idea is the following. The local bias only takes into account errors missed for variables involved in the edits failed for each record. However, since edits form a complex set of restrictions on the data, it is not uncommon that a record initially flagged for errors in one edit, say, are modified in a way that affects variables not involved in this edit. To understand this idea, an example may be helpful. Consider the following simple scenario of a business survey with only 5 questions (variables):

Q1 = Turnover (total)
Q2 = Total costs
Q3 = Employment costs
Q4 = Other costs
Q5 = VAT included in turnover reported in Q1

Suppose also that the following edits are applied to the data:

E_1: Q1 > Q2
E_2: Q2 = Q3 + Q4
E_3: 0.1 < Q5/Q1 < 0.3

Figure 1 contains an example of a record as it stands in two different moments during the survey process: at point of capture (raw) and at point of estimation (edited). The values of the five variables are provided, together with indicators of failing (1) or passing (0) each of the specified edits. Observe that the record initially fails only edit E_1, which involves the variables Q1 and Q2. However, when it was verified, the value of Q2 was changed, and as a consequence, because Q2 is related to Q3 and Q4 by another edit, the value of Q3 was also modified to enable the record to pass all edits, otherwise it would then fail E_2.

Figure 1 – Example of a record used to illustrate concept of global bias indicator

Version of the data	Q1	Q2	Q3	Q4	Q5	E_1	E_2	E_3
Raw	120	150	140	10	25	1	0	0
Edited	120	110	100	10	25	0	0	0

Suppose that we modified the edits such that:

E_1^*: 2Q1 > Q2

It is easy to verify that this record would now not fail any edits in the revised set, and hence would be 'saved' from entering the validation operation, but as a consequence would have all its errors 'missed', i.e., left undetected. Assuming that the changes previously made actually were corrections to errors found in the data, not editing this record would cause some bias for Q2 (which was changed and was involved in edit E_1), but using the concept of local bias, would not cause a bias for Q3 (which was changed but was <u>not involved</u> in edit E_1). However, this does not mean that there would not be a bias for Q3. Indeed, it is also easy to see that the bias for Q3 would be even larger than that for Q2, since a difference of 40 is larger if the denominator includes the weighted edited value of Q3 (100 for this record) than if it includes the weighted edited value of Q2 (110 for this record), other things being held constant.

In addition, for this very naïve example, because the number of variables involved in the edits is only 5, the average local bias would be the bias for Q2 divided by 5, whereas the average global bias would have included the sums of the biases for Q2 and Q3 divided by 5, and would be at least approximately twice as big as the average local bias.

Hence our concept of '*global bias*' provides a more comprehensive measure of the impact of changing the edits, by considering any changes made to a record that is missed as a result of changing the edits, even when they affect variables which were not active in the original edits failing the record.

One other small difficulty that we had to deal with when implementing these ideas in developing a revised SNOWDON tool was due to the fact that the edits used on ABI2 often include derived variables as part of their formulation or specification. If this is not a problem in terms of record detection, it creates some difficulties because derived variables are irrelevant for error localization. The respondents can only make an error in the original variables contained in a questionnaire, and the edit reviewers should only make changes to the values of the original variables in the record. In addition, many of the derived variables are not linear functions of the original variables. In fact, very often they are ratios or relative changes computed from original variables or their values in previous survey rounds. In such cases, the bias measures (either local or global) for such derived variables are irrelevant or meaningless, since we do not plan to estimate their population totals, and the bias indicators are simply expected changes in the estimates of the population totals when some of the errors formerly detected and corrected would be missed if we changed the edits in some specified manner.

To deal with this difficulty, we computed the summaries (C6), (C7), (C8) and (C9) only from bias (local or global) arising out of the *original variables*, excluding any *derived variables* from these calculations. In the future, it may be desirable to include some of these derived variables which are linear combinations of the original variables in these summaries, such as is the case of the GVA (Gross Value Added), since for this variable the estimates of total are published.

We note also that the average summaries in (C6) and (C8) are simple unweighted averages of the relative bias (local or global) for the various variables. If required, other summaries could be used, say by replacing these by weighted averages, with weights that reflect the relative importance of the various variables in the set of variables involved in the edits. As an example, the average relative global bias could be estimated by:

$$RGB_{\alpha} = \sum_{j=1}^{p} \alpha_j \hat{B}_j^G \Big/ \sum_{j=1}^{p} \alpha_j \qquad (4)$$

where $\alpha_j \geq 0$ is the importance weight assigned to variable j, j=1,…,p.

To provide a 'proof of concept' of the usefulness of this revised and extended set of indicators of edit performance and on the impact of possible changes to a set of edits, a SAS macro named **SNOWDON-X** was developed to implement computation of all the proposed indicators of edit performance. This is currently being tested by staff in the Methodology Directorate Process Editing and Imputation team. The code of this macro is available from the authors on request.

The essential input information for this macro includes the following:

(a) Two versions of the survey data for a given period– *raw* and *edited*; these data sets should have identical structure, namely rectangular files in SAS format, with one row for each responding unit, and containing at least the following sets of variables – record identification, frame related variables (classification, size band, design weight, calibration weight, final weight) and all relevant survey questions (both original and derived);

(b) Two versions of the set of edits – *original* and *revised*; the edits should be specified in the form of '*IF expression THEN action*'; the *expression* must be a valid SAS data step expression, involving original or derived variables, and logical operators; *expression* should indicate a condition when a record FAILS an edit; *action* sets an edit flag equal to one for each of the existing edits;

(c) Two versions of a table providing the indicators of which variables are active in each of the edits; each of these tables should be a SAS data set with rows corresponding the edits and columns corresponding to the variables; one table should correspond to the *original edits*, and the other to the *revised edits*; it is possible that in some applications both tables are the same; this would happen when changes to the edits do not alter the set of variables involved in each of the edits.

The outputs of this macro include:

(a) a table containing the general edit performance indicators associated with one set of edits and its revised version;

(b) a set of performance indicators for each edit in the set E;

(c) a set of quality indicators for each variable;

(d) a dataset containing the raw values, edited values, and revised values for all records and variables, which can be used to compute further indicators of the impact of changing the edits, if required.

Some initial experimentation with this program suggests that it is very fast to execute, dealing with data from a large sector such as Production and Construction short questionnaire in under one minute in a standard office PC.

5. Application Of Snowdon-X To ABI2

In this section an application of SNOWDON-X to the ABI2 survey data is described. First we considered only data obtained using the short questionnaire in the catering sector. The reason for initially choosing this sector and questionnaire was operational: the corresponding data and edits have been previously processed and organised in the context of the preparation of the first report on ABI2 editing review - see Silva (2007). In addition, the catering short questionnaire provides an example of a sector with some high edit failure rates coupled with low hit rates, hence where the potential for savings from edit revision may be substantial. The data considered for this analysis were the raw data after application of the automatic editing procedure which aimed to detect and correct for the so-called ‘thousand pound errors’ - see Silva (2007) for a discussion.

Following Silva (2007), the failure rates for each original edit applied to the catering short questionnaire were examined for three years: 2003, 2004 and 2005. Table 5 presents the edit failure counts and rates for the edits applied in 2004 by descending number of failures. Two immediate conclusions emerge. First, there are two edits (3102 and 3100) which have much higher failure rates than all the others, with failure rates in excess of 37% of the records each. Hence if any substantial savings are to be made, these edits must be targeted for change. Second, there are two edits (1151 and 1152) which flagged no records in 2004. Hence under similar conditions for future surveys, changes to these edits would not be expected to produce any savings. In fact, it might be argued that they are perhaps unnecessary. Similar patterns were noted while examining data from other survey periods (2003 and 2005) as well as for some other sectors, namely a few edits present high failure rates, and most other edits have small failure rates.

In view of these findings, the following procedure for reviewing the edits in a set is proposed.

1. First sort the edits in descending order according to their failure rates in a recent survey period, i.e., edits with higher failure rates are on top of the list. Visual displays like Pareto charts of the edit failure rates can be used to highlight the top failing edits, as noted by Silva (2007).
2. Locate and revise the first edit in the set, either by filtering, changing its tolerance parameters or its definition, or even by discarding it.
3. Re-calculate the edit failure rates, the expected savings, and the expected bias resulting from revision of this edit.
4. If the savings and quality statistics are not satisfactory, refine your tolerance parameters, or the edit definition. When you are satisfied with the outcome of the revision for this first edit, proceed to revise the next edit down the list.
5. Stop when the expected impact of further revisions in terms of savings is minor, or when the expected impact upon the quality indicators is large or reaches a specified threshold.

By proceeding in this fashion the analyst will gain an insight into the potential impact of the changes made to each of the revised edits. Hopefully the process will stop after the first few most failed edits have been revised. Hence it should not take a long time to do for each set of edits and data under consideration.

To test the proposed approach, it was applied to review edits for the short questionnaires used in the ABI2, aiming to replace those currently used in the editing for the survey of reference year 2007 (questionnaires currently in the field for collection). Data and edits from 2005 were used for the analysis of edit performance and as the basis for edit revision. Short questionnaires were chosen for two reasons: first they are more numerous, responding for around 1/3 of the questionnaires in the ABI2 survey each year. Second, they are used to obtain information from the smaller businesses, which often enter the survey for a single round, and are thus more likely to make mistakes in filling out the questionnaires. Third, these are the simplest questionnaires in the ABI2 data collection, hence they have smaller sets of edits.

In what follows, a summary of the results of this exercise are presented. The complete and detailed report describing the exercise is Bucknall and Zong (2008). Data were examined separately for seven sectors of the economy. Table 6 in the Appendix provides some summary measures describing the results of the application. The aim of this exercise was to enable survey managers to obtain some modest reduction in the editing effort without a major overhaul of the editing system, hence the cautious approach of changing only a few edits per sector, and even there, only making changes that would not lead to important bias in the survey estimates. The goal was to try and achieve a 5% reduction in the editing workload. The revisions proposed would achieve an estimated 6% reduction, while keeping bias under control for all sectors (the largest estimated absolute relative bias was 0.65% for the catering sector). Figure 2 in the Appendix illustrates how the edits would perform for the catering sector before and after revision.

Table 7 in the Appendix presents an example of the summary indicators produced and presented for each run of the SNOWDON-X macro, after some revision was made to the original edits. By looking at tables like these, the analyst reviewing the edits can keep an eye on the expected impact of changes to the edits, and hence make informed decisions on how to proceed at each step. Detailed description of all the changes proposed to the edits and the outcomes of the exercise are beyond the scope of this paper, but those interested in such details should consult Bucknall and Zong (2008).

6. Conclusion And Future Developments

This paper described the SNOWDON-X tool developed to provide a comprehensive set of edit performance indicators which can be readily used to assist survey managers and designers to both assess the potential impact of edit revisions, as well as to monitor edit performance in cases where the edits are not modified from one survey period to the next. SNOWDON-X should be a useful addition to the toolkit for those in charge of assessing edit performance in connection with business surveys, in the ONS and perhaps also elsewhere.

We illustrated how SNOWDON-X was used to assist in revising the edits for the ABI2. Our proposed approach includes guidelines on how one might use the indicators provided by SNOWDON-X tool to target edits for revision, and to investigate how each edit revision may impact the overall editing operation. Specific advice on how to revise each edit is beyond the scope of this paper, but the reader is encouraged to take a broad perspective and to consider the various options available, including the edit deletion option. The benefits of the expanded set of indicators provided by the SNOWDON-X tool were clearly demonstrated, as well as the ability of the tool to deal with a very complex survey as is the case of ABI2.

Use of the SNOWDON-X tool in some automatic fashion to try out varying values of the tolerance parameters, chosen without careful analysis of the edit specifications, is strongly discouraged. If this speeds up some computations, it does not lead to better understanding of the edits and consequently, no real improvement to the survey process can be expected.

We recognise that the estimation of impact of the revised set of edits is dependent on a strong assumption that the survey data used for the analysis is error free, when the edited version of the data is considered. This assumption often does not offer a good description of the situation in practice. However, even when the assumption is not true, the indicators provided by SNOWDON-X should still be useful for those seeking to make revisions to a set of edits. It is likely, however, that the impact of changes to the edits is not well predicted in such cases.

We also recognise that marginal edit revision in a given survey may not be the best approach. Some surveys might benefit from a more thorough revision of the whole editing approach, possibly replacing traditional methods with modern methods, including significance editing, automatic editing and imputation, macro-editing, etc. However, major revisions of the editing process of some surveys may take a long time to complete and are likely to be a very costly exercise. The SNOWDON-X tool might provide useful assistance when such major revisions are not immediately feasible, and only minor adjustments to the edits are an option. We stress that the edit designer or survey manager should not give up on the idea of in-depth analysis of the edits and their impact over the data, whatever the size of the redesign at hand. Quick fixes may sometimes prove costly at later stages of survey processing.

References

Al-Hamad, A., Martín, B., and Brown, G. (2006). SNOWDON: Tool for Assessing Impact of Changing Editing Rules on Cost & Quality. *Survey Methodology Bulletin* **60**, 26-35.

Bucknall, R. (2006). Impact of cutting validation in the ABI/2 (Catering Sector) - DRAFT. 15. Newport: Office for National Statistics.

Bucknall, R., and Zong, P. (2008). SNOWDON assessment of ABI/2 results. 23. Newport: Office for National Statistics.

Chambers, R. L. (2001). Evaluation criteria for statistical editing and imputation. In *National Statistics Methodology Series*, J. Charlton (ed), 41. Newport: Office for National Statistics.

Luzi, O., Di Zio, M., Guarnera, U., *et al.* (2007). Recommended practices for editing and imputation in cross-sectional business surveys - draft version 1.4. 73: ISTAT Instituto Nazionale di Statistica, CBS Statistics Netherlands, SFSO Swiss Federal Statistical Office.

Silva, P. L. d. N. (2007). Editing and imputation for the ABI/2 survey - interim report 1. 72. Southampton, UK: S3RI - Southampton Statistical Sciences Research Institute

Appendix – Tables and Figures

Table 1 – General descriptive indicators for the data set under analysis

Description	Definition	Equation
Number of Complete Records, i.e. records with no missing values	$R = \sum_{i \in r} r_i$	(A1)
Proportion of Complete Records	$\overline{R} = R / m$	(A2)
Number of Incomplete Records, i.e. records with at least one missing value	$\tilde{R} = \sum_{i \in r} (1 - r_i)$	(A3)
Proportion of records with at least one missing value	$1 - \overline{R} = \tilde{R} / m$	(A4)
Number of records failing at least one edit rule	$V_+ = \sum_{i \in r} v_i$	(A5)
Proportion of records failing at least one edit rule	$\overline{V}_+ = V_+ / m$	(A6)
Number of 'clean' records, i.e., records failing no edits	$m - V_+ = \sum_{i \in r} \tilde{v}_i = \sum_{i \in r} (1 - v_i)$	(A7)
Proportion of 'clean' records, i.e., records failing no edits	$1 - \overline{V}_+ = (m - V_+) / m$	(A8)
Number of records changed during editing	$G_+ = \sum_{i \in r} g_i$	(A9)
Proportion of records changed during editing	$\overline{G}_+ = G_+ / m$	(A10)
Number of records unchanged during editing	$m - G_+ = \sum_{i \in r} \tilde{g}_i = \sum_{i \in r} (1 - g_i)$	(A11)
Proportion of records unchanged during editing	$1 - \overline{G}_+ = (m - G_+) / m$	(A12)

Table 2 - Cross-classification of data value status and error flag for variable *j* in record *i*

Status of raw data item z_{ij}	**Error flag for raw data item** z_{ij}		**All records**
	Data item flagged $v_{ij} = 1$	**Data item not flagged** $v_{ij} = 0$	
Incorrect / changed $g_{ij} = 1$ or $z_{ij} \neq y_{ij}$	**Correct decision** $g_{ij} v_{ij} = 1$	**Missed error** $g_{ij}(1 - v_{ij}) = 1$	$G_j = \sum_{i \in r} g_{ij}$
Correct / unchanged $g_{ij} = 0$ or $z_{ij} = y_{ij}$	**False detection** $(1 - g_{ij}) v_{ij} = 1$	**Correct decision** $(1 - g_{ij})(1 - v_{ij}) = 1$	$\tilde{G}_j = m - G_j$
All records	$V_j = \sum_{i \in r} v_{ij}$	$m - V_j$	m

Table 3 – Indicators of edit performance prior to edit revision

Description	Definition	Equation
Total number of record failures for variable *j*	$V_j = \sum_{i \in r} v_{ij}$	(B1)
Failure rate for variable *j*, i.e. the proportion of records flagged by edits involving variable *j*	$\bar{V}_j = V_j / m$	(B2)
Total number of changes made to variable *j* – can also be interpreted as the total number of errors in variable *j*	$G_j = \sum_{i \in r} g_{ij}$	(B3)
Total number of records where variable *j* was not changed during validation	$\tilde{G}_j = \sum_{i \in r} \left(1 - g_{ij}\right) = m - G_j$	(B4)
Number of errors uncovered in variable *j*, i.e. the number of times that variable *j* was changed during validation when it was flagged by edits	$H_j = \sum_{i \in r} v_{ij} g_{ij}$	(B5)
Hit rate 1 for variable *j*, i.e. the proportion of times that variable *j* was changed during validation given that it was flagged by edits	$\bar{H}_j^1 = \sum_{i \in r} v_{ij} g_{ij} \Big/ \sum_{i \in r} v_{ij} = \dfrac{H_j}{V_j}$	(B6)
Hit rate 2 for variable *j*, i.e. the proportion of times that variable *j* was flagged by edits given that its value was changed during the editing	$\bar{H}_j^2 = \sum_{i \in r} v_{ij} g_{ij} \Big/ \sum_{i \in r} g_{ij} = \dfrac{H_j}{G_j}$	(B7)
Number of false hits for variable *j*, i.e. the number of times that variable *j* was not changed during validation despite being flagged by edits	$\tilde{H}_j = \sum_{i \in r} v_{ij} \left(1 - g_{ij}\right)$	(B8)
False hit rate 1 for variable *j*, i.e. the proportion of times that variable *j* was <u>not</u> changed after validation given that it was flagged by edits	$\bar{F}_j^1 = \dfrac{\sum_{i \in r} v_{ij}\left(1 - g_{ij}\right)}{\sum_{i \in r} v_{ij}} = \dfrac{\tilde{H}_j}{V_j}$	(B9)
False hit rate 2 for variable *j*, i.e. the proportion of times that variable *j* was flagged by edits given that it was <u>not</u> changed after validation	$\bar{F}_j^2 = \dfrac{\sum_{i \in r} v_{ij}\left(1 - g_{ij}\right)}{\sum_{i \in r} \left(1 - g_{ij}\right)} = \dfrac{\tilde{H}_j}{\tilde{G}_j}$	(B10)
Total number of edit failures for edit E_k	$e_{+k} = \sum_{i \in r} e_{ik}$	(B11)
Failure rate for edit E_k	$\bar{e}_k = e_{+k} / m$	(B12)
Total number of errors in records under original edits	$G = \sum_j G_j$	(B13)
Total number of false errors in records under original edits	$\tilde{H} = \sum_j \tilde{H}_j$	(B14)

Description	Definition	Equation
Overall hit rate 1, i.e. the proportion of times that fields were changed during validation when flagged by edits	$\overline{H}^1 = \sum_{i\in r}\sum_j v_{ij} g_{ij} \Big/ \sum_{i\in r}\sum_j v_{ij} = \sum_j H_j \Big/ \sum_j V_j$	(B15)
Overall hit rate 2, i.e. the proportion of times that fields were flagged by edits given that their value was changed during the editing	$\overline{H}^2 = \sum_{i\in r}\sum_j v_{ij} g_{ij} \Big/ \sum_{i\in r}\sum_j g_{ij} = \sum_j H_j \Big/ \sum_j G_j$	(B16)
Overall false hit rate 1 = proportion of times that fields were not changed after validation when flagged by edits	$\overline{F}^1 = \dfrac{\sum_{i\in r}\sum_j v_{ij}\left(1-g_{ij}\right)}{\sum_{i\in r}\sum_j v_{ij}} = \dfrac{\sum_j \tilde{H}_j}{\sum_j V_j}$	(B17)
Overall false hit rate 2 = proportion of times that fields were flagged by edits but were <u>not</u> changed validation	$\overline{F}^2 = \dfrac{\sum_{i\in r} v_{ij}\left(1-g_{ij}\right)}{\sum_{i\in r}\left(1-g_{ij}\right)} = \dfrac{\tilde{H}_j}{\tilde{G}_j}$	(B18)

Table 4 – Indicators of impact due to edit rule revisions

Description	Definition	Equation
Savings in number of records to be edited (total savings expected)	$\text{Savings} = V_+ - V_+^*$	(C1)
Total number of missed errors	$M = \sum_{i\in r}\sum_j \left(v_{ij} - v_{ij}^*\right) g_{ij}$	(C2)
Missed error rate	$\overline{M} = \sum_{i\in r}\sum_j \left(v_{ij} - v_{ij}^*\right) g_{ij} \Big/ \sum_{i\in r}\sum_j v_{ij}$	(C3)
Relative absolute *local bias* for variable *j*	$\hat{B}_j^L = \dfrac{\sum_{i\in r} w_i v_i (1-v_i^*) v_{ij} (1-v_{ij}^*) \left\| z_{ij} - y_{ij} \right\|}{\sum_{i\in r} w_i y_{ij}}$	(C4)
Relative absolute *global bias* for variable *j*	$\hat{B}_j^G = \dfrac{\sum_{i\in r} w_i v_i (1-v_i^*) \left\| z_{ij} - y_{ij} \right\|}{\sum_{i\in r} w_i y_{ij}}$	(C5)
Average relative absolute local bias for all original variables involved in edits	$RLB = \dfrac{1}{p}\sum_{j=1}^{p} \hat{B}_j^L$	(C6)
Maximum relative absolute local bias for all original variables involved in edits	$MLB = \max_j \left(\hat{B}_j^L\right)$	(C7)
Average relative absolute global bias for all original variables involved in edits	$RGB = \dfrac{1}{p}\sum_{j=1}^{p} \hat{B}_j^G$	(C8)
Maximum relative absolute global bias for all original variables involved in edits	$MGB = \max_j \left(\hat{B}_j^G\right)$	(C9)

Table 5 – Edit failure indicators for catering, short questionnaire, 2004

Edit Code	Number of records failing edit	Edit failure rate (%)
3100	669	38.5
3102	648	37.3
1131	281	16.2
1123	165	9.5
3101	162	9.3
1172	120	6.9
1113	65	3.7
1111	64	3.7
1133	56	3.2
1141	52	3.0
1134	48	2.8
1112	21	1.2
1143	14	0.8
1124	8	0.5
1153	6	0.3
1154	6	0.3
1125	4	0.2
1151	0	0.0
1152	0	0.0

Table 6 – Summary of selected indicators for edit revision for ABI2 – short questionnaire

Sector	# of short questionnaire records	% of records in sector	# of edits changed	# of records failing edits	Savings	Maximum absolute relative global bias
Catering	1,712	66%	3	606	58	0.65%
Retail	3,809	68%	3	3,107	173	0.21%
Motor Trades	1,716	72%	3	810	90	0.31%
Service Trades	5,108	27%	2	1,162	67	0.04%
Wholesale	3,471	62%	3	1,815	110	0.29%
Property	1,062	70%	1	363	13	0.05%
Production & Construction	5,826	39%	1	2,348	106	0.53%
Overall	22,704	-	16	10,211	617	-

Table 7 – Edit performance indicators for catering, short questionnaire, 2005, after revising three edits:

Description of indicator	Values
Number of records in survey dataset	1,712
Total number of variables involved in edits	23
Number of original variables involved in edits	10
Number of edits applied to survey records	18
Number of fatal edits in edit set	1
Number of query edits in edit set	17
Number of incomplete records	0
Proportion of incomplete records (%)	0.00
Number of records failing at least one edit rule - original edits	606
Proportion of records failing at least one edit rule - original edits (%)	35.40
Number of records failing at least one edit rule - revised edits	548
Proportion of records failing at least one edit rule - revised edits (%)	32.01
Savings (records not sent to validation)	58
Total number of missed errors	136
Missed error rate (%)	3.59
Mean relative absolute bias - Global (%)	0.14
Maximum relative absolute bias - Global (%)	0.65
Mean relative absolute bias - Local (%)	0.12
Maximum relative absolute bias - Local (%)	0.65
Overal hit rate - original edits (%)	32.22
Overal hit rate - revised edits (%)	33.76
Overal false hit rate - original edits (%)	67.78
Overal false hit rate - revised edits (%)	66.24

Figure 2 – Illustration of edit performance indicators obtained using SNOWDON-X Catering – short questionnaire – 2005