WP.7
ENGLISH ONLY

UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Work Session on Statistical Data Editing
(Bonn, Germany, 25-27 September 2006)

Topic (i): Editing nearer the source

TOOL FOR ASSESSING IMPACT OF CHANGING EDITING RULES ON COST & QUALITY

Supporting Paper

Prepared by Alaa Al-Hamad, Begoña Martín, Gary Brown, Office for National Statistics, U.K.[1]

I. INTRODUCTION

1. In the late 1990s, a large efficiency programme was initiated by the Office for National Statistics (ONS) in the United kingdom (UK) with the vision of increasing the efficiency of data processing in order to:

(a) meet financial pressures; and
(b) generate savings for reinvestment in improved outputs.

2. Evidence from the literature suggests that one of the most costly components of survey processing in any national statistical institute (NSI) is data cleaning. This is because to fulfil their role successfully, most NSIs are under pressure to produce high-quality data in order to maintain the public confidence. The ONS is no different from other NSIs in this respect, so business surveys within the organisation tend to have high data cleaning costs that consume almost half of the total business survey budget.

3. A major function that consumes most of these resources/costs is data editing. In practice this consists of two stages.

(a) Error localisation/detection, mainly using what is referred to as editing rules.
(b) Error verification/correction, mainly by re-contacting the respondent.

4. Therefore when funding is reduced or efficiency savings are required, the editing process tends to be one of the first targets. An easy way to make quick savings, which is often used by survey managers, is to 'loosen' editing rules so that fewer errors are identified. However, making these kinds of decisions based just on judgement raises the question of where to stop, and what the impact will be on output quality.

5. The question discussed in this paper is how to assess the impact of reducing costs on output quality. We propose a decision support tool that will help survey managers evaluate what savings can be achieved, at what cost to output quality, across many alternative permutations of editing rule parameters. The outcome will be minimum quality loss for defined savings, or maximum savings for defined quality loss.

[1]Office for National Statistics, Cardiff Road, Newport, Gwent, NP10 8XG, UK
(Alaa.Al-Hamad@ons.gov.uk, Begona.Martin@ons.gov.uk, Gary.Brown@ons.gov.uk)

II. BACKGROUND, AND EDITING ISSUES

A. Data Cleaning in ONS

6. In the ONS the data cleaning model is very complicated. This is due to the variety of data collection modes. Data arrive into the office through a variety of modes: paper questionnaires, Telephone Data Entry (TDE), FAX, by telephone, and web collection through the internet. Figure 1 presents a simplified illustration of the data cleaning process, using the paper questionnaire mode as an example.

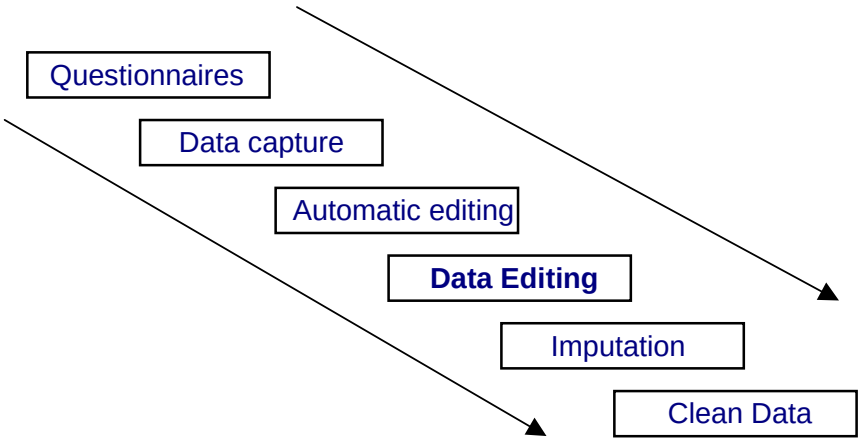

Figure 1: Data cleaning in ONS

- *Data capture*

7. When paper questionnaires arrive, the written responses they contain are captured electronically using scanning (mainly). Responses that cannot be scanned accurately (according to user specified confidence intervals) are subject to visual confirmation and data input.

- *Automatic editing*

8. Despite using best practice techniques in the design of questionnaires for business surveys, errors still occur as respondents make mistakes. These include systematic errors such as quoting in pounds instead of in thousands of pounds, and totalling (i.e. basic addition) errors. Many ONS business surveys automatically identify and correct these two systematic errors.

- *Data editing*

9. Data editing in the ONS is defined as an activity aimed at detecting and correcting errors in data. This involves detecting responses that are unlikely to be correct based on editing rules. Editing rules usually compare returned or derived data with past data, or with constants.

10. Data that fail editing rules are checked – and either verified as correct or corrected. This decision is made on the basis of auxiliary data, the scanned images of the questionnaires, prior knowledge, or by recontacting the respondent. However, recontact is the main method. For some surveys, selective editing prioritises editing rule failures (by considering their expected impact on final estimates). Thus selective editing could be used prior to checking, but this will not be explored by this paper.

- *Imputation*

11. After editing is complete, any missing or unusable data are imputed for (using a variety of methods). At this stage the data are clean and ready for analysis.

12. From this description of the data cleaning process, it is clear that editing is the most resource intensive/costly component. What needs to be borne in mind is this process is repeated for approximately 60 business survey of different sizes and complexities.

B. Error Detection Rules

- *Types of rules*

13. Editing rules usually take the form of simple mathematical, often linear, equations, with the majority having parameters that define the acceptable region for a response. For the purpose of this paper these rules can be classified as follows.

(a) Checking for missing values and the presence of data, i.e. non-response
(b) Checking for logical errors/inconsistencies, i.e. negative employment or whether responses in the same questionnaire are inconsistent
(c) Checking for statistical/value errors, i.e. the response falls outside likely bounds, so is suspected to be in error

14. The third and last type of rule, generally governed by parameters identifying the acceptable region, will be the focus of our attention.

- *Outcome of editing process*

15. Before the editing rules are applied, the raw data are either in error or not. The aim of the rules is to determine which responses are suspect, and which are non-suspect. Therefore the scenarios, if the rules worked perfectly, are as follows.

- A. Data with error (bad data) would fail
- B. Data without error (good data) would pass

16. However, as illustrated in Figure 2, two other less appealing outcomes also occur.

- C. Some good data fail
- D. Some bad data pass

	Pass validation rules	Fail validation rules
Error in data	D - errors missed	A - errors detected
No error in data	B - saved effort	C - wasted effort

Figure 2: Error detection outcomes

17. Therefore, in order to achieve an ideal detection rule, the following two objectives should be met:

(a) maximise A and minimise C; and
(b) maximise B and minimise D

C. Rule Adjustment and Implications

- *Rule adjustment*

18. Due to the nature of value error rules, the parameters govern these rules determine which data fail and which pass. If parameters are changed, the numbers failing and passing change. Striking the right balance in setting these parameters is not always easy.

- *Implications*

19. If rule parameters are set too tight:
    - Increased response burden due to unnecessary recontact
    - Reduced data quality due to over-validation and associated biases
    - Increased cost in terms of staff and resources

20. If rule parameters are set too loose:
    - Uncorrected errors pass the rules
    - Reduced data quality as uncorrected errors will increase and persist
    - Costly in terms of ONS reputation

21. However, the latter action is less costly in terms of staff and resources if we have the right tool to help survey managers to decide how loose the rules can be made.

D. Saving vs Quality trade-off

22. An easy way for survey managers to make quick savings on editing is either to drop editing rules, or to loosen the parameters in value error rules - to let more data pass.

23. However the manager faces two dilemmas in this scenario:
(a) To what extent can these rules be loosened?
(b) What impact will this have on the quality of output estimates?

24. As mentioned earlier, reducing survey costs, by reducing editing failures, will have a negative impact on quality: fewer errors will be identified, so fewer will be corrected. The performance of editing should be measured not only in terms of the number, but also the size and the impact of these uncorrected errors. Currently the two most common performance measures in use at the ONS are the failure rate and the hit rate.
    - The number of editing failures, as a proportion of all data, defines the *failure rate* for a particular rule (or across all rules).
    - The *hit rate* is the proportion of editing failures that are actually found to be in error. This represents, to some extent, the efficiency of identification – a low hit rate means that many responses were unnecessarily edited.

25. Evidence from the literature, for example Tate et al. (2001), indicates that in the majority of ONS business surveys, the failure rate is in the range of 30% to 80%, whereas the hit rate is generally around 20%.

III. SOLUTION

A. Suitable Measurements

26. Defining suitable performance measures will enable survey managers to make informed decisions.

- *A measure of savings*

27. The number of records no longer failing editing rules will help managers determine the reduction in costs and time, i.e. resources.

- *A measure of errors missed*

28. The hit rate before and after parameters have been changed determines the number of errors missed. At this stage of the process, managers should identify (and possibly reconsider) those businesses in error that are no longer edited.

- *A measure of the impact on quality*

29. The hit rate delivers a raw count of errors missed, but does not determine their size or impact on the overall output quality. Therefore, in this paper we define an impact measure - the relative change (in the survey estimates). We consider this the key quality measure. The relative change RC_D in domain *D* is defined as:

$$RC_D = \frac{\left|\sum w\left(X_{Before} - X_{After}\right)\right|}{\sum w X_{Before}} * 100 \qquad (1)$$

where: X = response before and after parameter change; and
w = calibration weight for the respondent.

30. Equation (1) measures the weighted errors missed due to changes in editing rules, relative to the weighted estimates under full editing. With no errors missed, RC_D will be 0. For errors missed, X_{Before} will be the edited value and X_{After} the unedited value.

31. Ideally, quality measures should also encompass bias and variance, but these have not yet been fully developed. As variance is not always known for ONS business surveys, this poses an additional problem.

B. Decision support tool

33. Ideally, quality measurements should be incorporated in a dynamic decision support routine for editing rules parameters, applicable generically to all business surveys. This would enable managers to assess the savings and impact of different scenarios of changes to edit rule parameters. Specifically, the routine should ideally:

- offer a choice of different quality measurement criteria;
- consider all editing rules simultaneously;
- output proposed changes to parameters; and
- output saving and quality loss per changed rule and in total.

34. A fully fledged generic dynamic version of this routine is still under development. Instead a pragmatic routine, incorporating the same criteria, was recently implemented on a real ONS business survey.

IV. IMPLEMENTATION USING ONS BUSINESS SURVEY

A. Quarterly Stocks Inquiry

35. The Quarterly Stocks Inquiry (QSI) is a sample based survey covering all of the UK. There are approximately 1.05 million businesses in the 8 sectors covered by the QSI. The overall sample size each quarter is approximately 21,500, i.e. around 2% of the population. The response rate is 75%. The QSI collects data on the level of stocks at the beginning and end of each period (i.e. opening and closing stocks).

36. In terms of outputs, the primary use of the QSI is in estimation of changes of stock (and of work in progress) for the expenditure and income measures of Gross Domestic Product (GDP). In addition, the QSI is used in the compilation of the Index of Production (IOP) which forms part of the output measure of GDP.

37. The QSI has, on average, a failure rate of 50%: approximately 30% of these errors are in the Wholesale & Retail Sector (WS&R). This paper will therefore focus on testing proposed changes in editing rules for the WS&R, in the QSI.

B. Current and Proposed Rules

38. After elementary analysis of all rules and discussion with survey manager, three rules were identified as potential sources for savings in the WS&R sector.

- Rule 1: Relative change in opening stocks
- Rule 2: Relative change in closing stocks
- Rule 3: Comparison of current opening and previous closing stocks

39. Let K denote thousands of pounds, and $y_i^{op}(t)$ and $y_i^{cl}(t)$ denote opening and closing stocks (respectively) at time t for a given respondent i, then the equations for these rules are as follows.

Rule 1. If $y_i^{op}(t) > 199K$, then fail if $\left|\frac{y_i^{cl}(t) - y_i^{op}(t)}{y_i^{op}(t)}\right| * 100 > 40$ (2)

Rule 2. If $y_i^{cl}(t) > 199K$, then fail if $\left|\frac{y_i^{op}(t) - y_i^{cl}(t)}{y_i^{cl}(t)}\right| * 100 > 40$ (3)

Rule 3. If $y_i^{cl}(t-1)$ was returned, then fail if $\left|y_i^{cl}(t-1) - y_i^{op}(t)\right| > 5K$ (4)

40. The proposed rules are based on replacing the existing thresholds with 5 variable parameters. For example, in Rule 1, the threshold of 199K is replaced by *Gate1* and the tolerance of 40% by *Tolerance2*. A new threshold *Gate2* > *Gate1* is then introduced such that if $y_i^{op}(t)$ is greater than *Gate2, Tolerance2* is replaced by *Tolerance1*. Equation (5) illustrates the new Rule 1 resulting from these changes.

Rule 1*. If $Gate2 > y_i^{op}(t) > Gate1$, then fail if $\left|\frac{y_i^{cl}(t) - y_i^{op}(t)}{y_i^{op}(t)}\right| * 100 > Tolerance2$

(5)

If $y_i^{op}(t) > Gate2$, then fail if $\left|\frac{y_i^{cl}(t) - y_i^{op}(t)}{y_i^{op}(t)}\right| * 100 > Tolerance1$

41. The change to Rule 2 is the same, and for Rule 3 the 5K threshold is replaced by *Gate3*.

C. Simulation Study

42. We vary the parameters of Gates 1- 3 and Tolerances 1 and 2 simultaneously.

43. Simulations use quarterly data from 2003-05. For each quarter, the first step is to split the data into two groups: P_1 corresponding to those records that failed editing and changed value, and P_2 corresponding to those records that failed but did not change value.

44. The simulation study consists of running the old editing rules on the data to determine the number of records that belong to group P_1. Editing is then rerun using the proposed rules, and records in P_1 that do not fail the new editing are counted. This count represents the number of genuine errors that will be missed under the new rules. We call it *missed errors*.

45. Conversely, we run the old editing rules to determine the number of records belonging to group P_2. We then rerun the proposed editing rules and count the records from P_2 that no longer fail. This count represents the reduction in the number of unnecessary failures. We call it *unnecessary errors*.

46. These two measures define s*avings*, the reduction in total editing failures, i.e.

$$savings = n_{missed\ errors} + n_{unnecessary\ errors}. \tag{6}$$

47. Finally, based on formula (1), the relative change is given by equation (7), where $y_i^{(ch)}$ denotes the raw return for 'change in stocks' of respondent *i* and $\tilde{y}_i^{(ch)}$ denotes the edited value. We define 'change in stocks' as the difference between the closing and opening stocks for that period.

$$RC_{WS\&R} = 100 * \frac{\left| \sum_{S_{ED}} w_i (y_i^{(ch)} - \tilde{y}_i^{(ch)}) \right|}{\sum_{S_D} w_i \tilde{y}_i^{(ch)}} \tag{7}$$

48. This process is repeated for every possible permutation of parameter values in Table 1. At each repetition, savings, missed errors and relative change are calculated.

Gate 1	Gate 2	Tolerance1&2	Gate 3
250,300	950,1000	40 & 50	10,20,30,40
400,500	1500,2000	40 & 55	50,75,100,1301
600	2500,3000	40 & 60	50,175,200

Table 1: Parameter values tested

D. Results

49. Table 2 displays the savings, missed errors, and relative change for various sets of parameter values for *Gate1, Gate2,* and *Gate3*. (*Tolerance1* and *Tolerance2* had no impact.) The data used were quarter 3 of 2005, the latest set of data available. Other quarters showed similar results. The table has been sorted by relative change.

Rules			Routine Output		
Gate1	**Gate2**	**Gate3**	**Savings**	**Errors Missed**	**Relative Change (%)**
600	40	**10**	111	77	0.56
600	40	**50**	205	171	1
600	40	**40**	192	158	1.28
600	40	**20**	160	126	1.32
250	40	100	243	209	2.96
300	40	100	243	209	2.96
600	40	100	243	209	2.96
600	**30**	200	274	240	3.93
600	**40**	200	274	240	3.93
600	**50**	200	274	240	3.93

Table 2: Results for quarter 3 of 2005

50. In Table 2 we observe that the relative change increases with the savings. In addition, the only parameter causing changes in savings, missed errors and relative change is *Gate3* (in Rule 3). The results suggest that excluding Rules 1&2 from the editing system would not impact on the relative change and savings achieved.

51. By setting *Gate3* to 10 we achieved sensible relative change for all quarters (except one). As a by-product of our analysis, the volatility across quarters suggested to us that setting parameters by period would be recommended for seasonal surveys.

52. The ultimate aim of this decision support table is to provide the customer with a simple but effective tool to:

- calculate the impact on quality resultant from making changes to edit rules;
- identify the combination of parameters that produces the minimum quality loss whilst achieving a pre-set level of savings.

53. The decision about the acceptable level of quality loss would remain with managers.

54. For example, if the customer seeks the smallest loss of quality whilst achieving savings of over 100 edit rule failures, the optimum parameters would be *Gate1=600, Gate2=1000, Tolerance1=40, Tolerance2=50,* and *Gate3=10*. These parameters would achieve an average saving of 111 failures per period, of which 69% were genuine errors, and would only cause a relative change of 0.56% on the estimate of 'change in stocks' in the current period.

V. CONCLUSIONS AND FURTHER WORK

- *Conclusions*

55. Until now, managers have made changes to editing rules one at a time and without consideration of the impact the changes will have on the quality of the survey output.

56. In this paper we define a simple but effective decision support tool that quantifies the loss in quality resulting from making changes to rules in order to reduce editing costs. Survey managers (and potentially customers) are supplied with a table that shows savings alongside parameters (for these savings), numbers of errors that will be missed, and impact on final estimates. Sorting this table (by impact and savings) will identify the parameters that cause the minimum impact for the targeted savings.

- *Further work*

57. Further work is dominated by development of a more dynamic, generic, program that can be applied to a variety of business surveys, to solve the problem considered in this paper. This will be a significant advance for data cleaning in the ONS, and will have positive implications for efficiency and quality throughout the survey process.

58. Other elements of further work are:

- investigating varying the parameters by domains (for example, Standard Industrial Classification (SIC) and employment sizeband);
- enhancing the impact measure by introducing variance and bias elements;
- testing data driven editing methods, for example Hidiroglou-Berthelot (1986); and
- continuing work to improve other stages in the data cleaning process, as an alternative to editing.

References

Granquist, L. and Kovar, J. (1997), "Editing of Survey Data: How Much is Enough?", *Survey Measurement and Process Quality*, eds. L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, and D. Trewin, New York: Wiley, pp. 415-435.

Tate, P., Underwood, C., Thomas, P. and Small, C. (2001), "Challenges in Developing and Implementing New Data Editing Methods for Business Surveys". *Proceedings of Statistics Canada Symposium 2001, "Achieving Data Quality in a Statistical Agency: a Methodological Perspective"*.

Hidiroglou, M.A. and Berthelot, J.M. (1986), "Statistical Editing and Imputation for periodic business surveys", *Survey Methodology*, June 1986, Vol. 12, N1, pp. 73-83.