WP. 15
ENGLISH ONLY

UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Work Session on Statistical Data Editing
(Ljubljana, Slovenia, 9-11 May 2011)

Topic (iii): Macro editing methods

Improving efficiency by introducing macro editing in Statistics New Zealand business performance surveys

Invited Paper

Prepared by Tim Hawkes, Statistics New Zealand

I. Introduction

1. Recent quality reviews of two of Statistics New Zealand's business performance surveys recommended reducing the number of manual micro edits to cut processing costs. These reviews concluded that "there is not enough emphasis placed on the macro editing phase. This should provide the foundation upon which other editing processes are based to provide the analyst with some direction and a 'big picture' perspective." In addition, an internal staff survey found that team members spent about "two man years" on micro editing across all of the business performance and energy suite of surveys. In response to this, a new approach to editing was developed and implemented in five 2010 business performance surveys.

2. This paper details this new approach to editing. It also presents the results of a review of the new approach using recently developed quality indicators for editing and imputation (E&I), and anecdotal evidence of efficiency gains. The results show a clear reduction in the time spent reviewing micro edits, an increase in the time spent macro editing, and an increase in the range of macro checks used.

3. The Business Performance team operates a suite of surveys that measure business operations, research and development (R&D), supply and use of information and communication technologies (ICT), biosciences, the screen industry, and energy use. Five of the business performance surveys ran collections in 2010 using the new editing approach: the Energy Use Survey, Business Operations Survey, ICT Supply Survey, R&D Survey, and the Screen Industry Survey. The ICT Supply and Screen Industry Surveys are full-coverage surveys while the other three surveys are sample surveys. Further details about these surveys can be found on the Statistics New Zealand website (www.stats.govt.nz).

4. The Business Performance team is a group of analysts who are responsible for the end-to-end production of the surveys. Each survey has a lead analyst. The Business Performance team receives methodological support from the Statistical Methods unit.

II. Previous editing approach to business performance surveys

5. Before implementing the new approach, editing for business performance surveys generally consisted of a large amount of time spent on manual micro editing followed by a limited amount of macro checking.

6. Micro edits almost always fell into one of three categories:

- validity edits verify whether the responses are either valid, permissible, or missing, given expectations of the data
- consistency edits check whether the expected relationships between responses are respected
- statistical edits verify whether responses align with expectations, where those expectations are based on statistical analysis or treatment of respondent data.

7. Edits check rules were implemented in the survey production system, which is called Legolution. If any record failed the edit checks then it was alerted in Legolution. Almost always, these alerted records were manually reviewed; however, on a small number of occasions, automatic edit rules were set to automatically deal with some of the alerts.

8. Macro editing would generally only start once manual micro editing was complete, and usually at a time when publication deadlines were placing pressure on timeframes. As such, the range of macro editing methods employed was fairly limited and were not standard across surveys. The most common approach was to produce ranked lists of the largest contributors to key level estimates and movements. There was some checking of key estimates against expectations and some drilling down of totals to component variables. Likewise, there was some checking of large contributions from imputed data and sample errors; however, these were usually calculated very close to publication dates.

III. Review of editing approaches

9. Before establishing details for the new editing approach, some of the Business Performance analysts undertook an evaluation of existing macro editing practices in other Statistics NZ business surveys. They also reviewed materials from the Australian Bureau of Statistics about their strategies for editing.

10. From this review, new potential strategies for both micro and macro editing were identified. Some of these strategies, such as the use of the graphical analysis package, Graphical Outlier Verification Initial Selections (Grover), were dismissed due to concerns about the amount of work required to implement and maintain. Other strategies were seen as being potentially useful but not practical to implement in the short term. These included possible improvements to questionnaires, the use of data cubes, investigations into software, including BANFF and LogiPlus, and the use of alternative survey production systems. However, several strategies were recommended to achieve the Business Performance team's goal of a more efficient editing approach.

IV. Recommendations for the new editing approach

11. The main recommendations for changes to existing practice fell into either the use of automatic micro editing or macro editing. The aim of these recommendations was to reduce the amount of time spent manually micro editing which would allow more time for macro checking.

A. Automatic micro editing

12. The following edits were automated.

13. *Consistency edits.* Legolution automatically edited errors attributed to either components of a sum not adding up to a total, or errors arising from a respondent failing to follow a routing instruction (called routing errors), or a logical error.

14. Previously, manual intervention was required even if the difference between the sum of components of a total and the reported total is less than $1, which more often may be due to rounding. With automation, only differences that are greater than some threshold are flagged for manual review. The thresholds could be relative (eg 10 percent of the total), absolute (eg $1,000), or a combination of relative and absolute (eg greater than 10 percent of the total or $1,000). Actual thresholds used were based on subject matter expertise, although in some cases it was possible to use data from previous survey cycles. In general, a conservative threshold is originally chosen. The threshold will be reviewed

after a number of survey cycles. A similar approach is taken with regard to errors in percentages since the sum of percentages of components should equal 100 percent. Thus, if the sum of percentages is within, say 10 percent (ie the total is between 90 percent and 110 percent), then the component percentages are automatically prorated to sum to 100 percent. If the sum is within 1 percent (ie the total is between 99 percent and 101 percent), then the component percentages are accepted.

15. Most routing errors were automatically edited. Table 1 shows some examples of routing errors and the actions taken to resolve them.

Table 1

Routing errors and actions to resolve them	
A routing question	**Possible automatic edit options**
Both 'yes' and 'no' then the section is blank	Auto edit to 'no'
Both 'yes' and 'no' then the section is answered	Auto edit to 'yes'
'No' then the section is answered	Auto edit to 'yes' or blank out section
'Yes' then section is blank	Auto edit to 'no' or leave for imputation

16. *Validity edits*. Since the business performance surveys used mailed questionnaires, multiple choice questions where only one answer was expected accounted for a number of errors. An example is a respondent ticking off 'yes' and 'no' to a question when only one of these if a valid answer. In these cases, automatically editing such types of errors was maximised. If a valid answer cannot easily be determined, the response is wiped out and left to be imputed. Table 2 shows some examples of invalid responses and the actions taken to resolve them.

Table 2

Invalid responses and actions to resolve them	
Categorical question which should have only one answer	**Possible automatic edit options**
One answer is 'other' or 'none of the above'	Auto edit so 'other' option is blank
One answer is much more common/expected	Auto edit to remove the unlikely answer
A series (eg income bands) with multiple selected	Auto edit to take the highest option
If no clear 'correct' answer	Auto edit to blank and leave for imputation

B. Macro editing

17. Previously, the macro editing done by the Business Performance team was limited because the majority of their processing time was spent dealing with all of the alerted micro edits. This meant that if any suspicious estimates were found then an investigation would put final publication dates at risk. Increasing the amount of automatic micro editing is intended to allow more time for a wider range of macro editing.

18. Macro editing should flow through to the analysis report for the clearance process. There should probably be no analysis done for the analysis report that is new. It should have all been done in macro editing with sufficient time to change the data, or if needed, to do special treatment (weight adjustment of extreme values) then repeat the analysis. In other words, the process of macro editing is iterative, leading to a final iteration where everything is explained and acceptable, no further changes to the data are needed, and the analyst is ready to write their analysis and get approval to release and publish.

19. In summary, the overall aim for macro editing is to ensure that our outputs are plausible. This is achieved by identifying possible influential errors in the data, determining further micro editing priorities, analysing data movements, and identifying influential (but correct) values for possible special treatment.

20. There were six main recommendations for an improved approach to macro editing.

Calculating estimates during the data collection process. The recommendation was to produce initial estimates once a survey has achieved 50–60 percent of its target response rate. This allowed early detection of influential observations which may need to be followed up. An issue that became apparent was insufficient responses for some imputation or weighting cells. Estimates from these cells were closely monitored as additional responses are collected. Another benefit of starting the estimation process early is that any problems with the estimation systems can be identified and resolved well before the production of the final estimates.

Top-down editing. This involved drilling down from the over-all total to look for unusual components. For example, a total economy estimate could first be checked at industry level, then at stratum level before checking microdata. It is possible that doing so may produce evidence that actually supports the estimate and suggests that expectations may have been wrong.

Use of sampling errors to identify suspicious estimates. All of the sample surveys in the business performance suite of surveys have expected sample errors from the sample design process. Final sample error estimates can be compared with expectations, along with sample errors from the previous year's surveys to help identify any suspicious estimates. Unusual responses that have an effect on sample errors may also have an effect on estimates. These should be checked and if correct, then possibly put forward as a candidate for special treatment.

Top contributor method. The three previous methods help identify suspicious estimates either at the overall level or at a lower level after drilling down. The next step is to check for unusual changes in final estimates. One way of doing this is to produce ranked lists of the top contributors to their level estimates, movements, or standard errors of estimates. The ranked lists of the top contributors are good for identifying responses that are too big, so that these can be either corrected or noted as candidates for special treatment. However it does not identify responses that are too small.

Ranking the top contributors to movements in estimates can help identify responses that may be either too small or too big. Whether or not all of the changes are in the same direction also helps with an understanding of what is driving a movement. These lists can be produced a number of ways depending on which weights are used.

It can also be useful to look at the biggest contributors to the sampling error of an estimate although this can be a bit trickier to produce and interpret.

Graphical analysis. The use of a box plot or a scatter plot can perform macro or micro editing. These plots have limitations, such as hidden data when plotting a large number of records, or detecting anomalies in more than two dimensions. However, problems are easier to see in graphs than in tables and they can also be helpful when originally defining automatic micro edit rules.

Processing checks. While all macro checks are useful for confirming that processes have run correctly, there are some specific checks that are essential for ensuring that the original data has been processed as intended. There are several types of checks and the choice of what checks to do will depend on the survey and the processing methods used. Some of the main examples are checks of: response rates (overall and by stratum), frequencies (overall and by subgroups), imputation ratios, and the range and sum of weights.

V. Implementing the new approach

21. Once the recommendations for automatic micro editing and macro editing had been identified, a training course was prepared by Statistical Methods for the analysts in the Business Performance team. This training covered the purpose of micro and macro editing and contrasted their current editing practice with the proposed approach. The training was interactive and the analysts were encouraged to come up with specific examples of how the recommendations could be incorporated in their relevant surveys. For example, they were asked to identify questions that were suitable to be edited using automatic consistency edits.

22. Beyond the training, no standard guidance was given as to the details of the new approach. It was up to the lead analysts to implement the new automatic edits in Legolution and to determine exactly what the macro editing phase would consist of. However, methodologists were available to provide any advice as necessary but generally the lead analysts were able to implement the new approach fairly independently.

23. Perhaps the only additional role that methodologists had in the implementation of the new approach was the development of a standard survey-monitoring template that contained the suggested processing checks. The template is designed to capture answers to all of the 'monitoring questions' that the Business Performance analysts must check to ensure that all of the statistical processes have run as specified, right from population and sample selection through to the confidentialisation of final outputs.

VI. Plan to review the new approach

24. Given that the recommendation for the new approach to editing was (for some surveys more than others) quite different from previous practice, it was deemed important for methodologists to conduct a review of how successfully the new editing approach had been implemented.

25. The review was also important as many of the recommendations contained subjective aspects. Decisions, such as those about thresholds or how many contributors to review for ranked lists, were basically left to the analyst's judgement. The approach for the first cycle of the new editing approach was to be fairly conservative with the subjective decisions with the possibility of 'tightening up' in future years; for example, tightening the thresholds, adding more automatic editing rules, or reviewing less contributors in ranked lists. The plan for the review was to try and incorporate both quantitative and qualitative feedback.

A. Quality indicators

26. Statistics NZ has recently developed a framework of E&I quality indicators. It has been developed from a collation of quality indicators that were sourced predominantly from national statistics organisations in Europe. The chosen quality indicators have been structured into four sections: monitoring and tuning indicators (for the subject matter area), publication indicators (for the user), supporting indicators (collection metrics, operational resources, and system quality) and indicators for testing (for the methodologist).

27. The most relevant indicators for the review of the new editing approach in business performance surveys are the monitoring and tuning indicators, specifically six measures in the editing quality subsection, and some of the indicators from the supporting indicators section.

Table 3 shows the relevant indicators from Statistics NZ's framework of E&I quality indicators.

Indicators from Statistics NZ's framework of editing and imputation (E&I) quality indicators	
Monitoring and tuning indicators	**Description**
Edit failure rate	Measures the proportion of edit failures out of the total number of records, overall and by edit rule
Average number of edit failures per unit	This is a measure of the average number of edit failures per unit, averaged over the units that have at least one edit failure
Clerical hit rate	Measures the proportion of values changed out of the total number of values clerically examined in the editing process
Clerical edits	Indicates how many units were clerically edited
Respondent re-contacts	Measures how many respondents were recontacted in the editing process
Error rate by error cause	Measures the proportion of values with a particular error cause out of all values clerically examined
Supporting indicators:	**Description:**
Duration of editing and imputation	Measures the elapsed time between the start and end of E&I, broken down by phase of the E&I process

Resource usage (staff time)	Measures the staff time used in E&I broken down by role and skill level and by phase of the E&I process

28. A subsequent case study, however, in which the quality indicators were applied to the 2009 Energy Use Survey, exposed some difficulties in compiling all of the data required to correctly calculate the indicators. In particular, some editing flags were either unavailable, unable to be extracted from Legolution, or were of questionable quality. As such it is not currently possible to calculate all of the above indicators for all five business performance surveys. In some cases only estimates of the actual value were possible.

B. Analyst feedback

29. A series of meetings were held with each of the lead analysts to establish the degree to which the recommendations had been implemented, their views on how successful the new approach was, any challenges they faced implementing the recommendations, and any thoughts they had on potential improvements to the editing approach in future years.

VII. Findings from the evaluation

30. Findings from the review of the implementation of the new editing approach in the five business performance surveys run in 2010, were as expected, generally positive. However, there were challenges faced and some areas for improvement were identified. All five of the surveys increased their use of automatic micro edits, which saw a reduction in the amount of manual micro editing done. Also, all five surveys had more thorough macro checks. Generally, the main goal of doing less manual micro editing and more macro editing was achieved although to varying degrees.

31. The R&D Survey delayed their micro editing phase because more time was needed to set up the automatic micro edits. Despite this, the team was still able to finish the micro editing phase earlier than in previous years and therefore had more time for macro editing. Since they will be able to reuse the automatic micro edits in future years, they expect to be able to free up even more time for macro editing.

32. The ICT Supply Survey had a change in methodology for dealing with non-response this year and the implementation used up much of the time saved in the micro editing phase. Similarly, the Energy Use Survey had delays before starting micro editing, which meant that the macro editing phase was not as long as hoped. However, both of these surveys still did more macro editing than in previous cycles.

33. The new editing approach worked well for the Business Operations Survey and the Screen Industry Survey in that the time spent micro editing was reduced and they were able to spend more time getting a better understanding of the data during the macro editing phase.

A. Automatic micro editing

34. All five surveys were able to reduce the time spent manually micro editing, which was not only beneficial in allowing more time for macro editing, but also was popular with the analysts who have, in previous years, not particularly enjoyed spending up to two months reviewing thousands of alerts. The ability to introduce more automatic editing varied from survey to survey and obviously depended on the questionnaire content.

35. For some surveys, implementing the new approach also served as a chance to look for additional questions for which new edit checks, as well as possibly automatic edit rules, could be added. This quality improvement meant the total number of edits (manual or automatic) actually increased in the R&D Survey, despite a large reduction in the number of manual edits.

36. The Energy Use Survey made an additional efficiency gain by removing an edit check that alerted very large values beyond a certain value. Previously, all of these were manually checked, even if

there was no evidence of actual errors. They removed this edit because they were confident that they would pick up any problems through the macro editing phase.

37. *Consistency edits.* All three of the surveys that had suitable questions successfully implemented automatic consistency edits. The Screen Industry Survey analysts found the writing of edit rules for routing questions especially complicated because there were several interactions between some of the questions. For example, it was not just a case of checking whether components summed to a total, but the components may have been totals themselves in breakdowns later in the questionnaire. This meant the ordering of the edit rules became very important.

38. The ICT Supply Survey implemented consistency edits for sums of components and the analysts felt as though they had chosen appropriate thresholds, as they had tested these thresholds on historical data. Likewise, the R&D Survey analysts believed they had chosen appropriate thresholds as, of the roughly 200 errors that fell outside of the thresholds, many of those were 'oddities' that were not edited in the same way as the automatic editing rule would have edited them. Several of these cases were addressed by calling back the respondent. In implementing the consistency edits, they actually checked their choices for thresholds with some of their external stakeholders.

39. The Energy Use Survey dealt with percentage questions (where the components should add up to 100 percent) with a consistency edit. They felt as though their choice of a threshold was about right as most differences tended to be either very small (for example, the components summed close to, but not exactly 100 percent) or fairly significant, so it was easy to choose a threshold that easily sorted these two types of errors. They chose to automatically accept these small differences rather than automatically edit them (to make them exactly equal 100 percent) as this was much simpler to implement. This meant that when the percentages were applied to magnitude values they didn't get perfect additivity of the totals, although the differences were insignificant, as expected.

40. *Validity edits.* The Business Operations Survey and the Energy Use Survey were the main users of validity edits. The Business Operations Survey already had several validity edits but the fixes were not previously automated. They found it to be a fairly straightforward process to write rules to automatically treat these errors (most of which were cases where more than one option had been selected when only one choice was asked for). In many cases, they actually just automatically accepted errors as they deemed them insignificant in final estimates.

B. Macro editing

41. All five surveys realised the benefits of an increased amount and wider range of macro editing, both in terms of identifying significant errors and gaining a more detailed understanding of the data. Other benefits were unforeseen, such as improving their ability to run the estimation programmes and understanding how features such as extreme weights can effect estimation.

42. The better understanding of the data gained though more macro editing for all surveys also helped in the final production stage. This meant that final release dates have not been put under pressure, even if the close-off on the processing was later than planned.

43. They were sometimes still slightly uneasy about the fact that it can be a subjective process with questions such as 'how can we be certain whether it is significant or not?' They also mentioned that the process could be improved by standardisation (in terms of exactly what macro checks they should do) and automation to make the process even more efficient.

44. *Calculating estimates during the data collection process.* As mentioned above, it was recommended to begin producing initial estimates at around 50–60 percent response (final expected response rates are 75–85 percent). As this was a new approach, it was unclear exactly at which points the response rate would be high enough to draw useful inferences from the early estimates. In reality it will always vary, depending on factors such as how many keys units have responded, and whether there are sufficient responses in each weighting/imputation cell.

45. The Business Operations Survey produced their initial estimates when the response rate was only 40 percent. These estimates produced some unusual results; however, early estimates were run without imputation, and all of the unusual estimates were explained by the lack of imputation. Subsequent iterations of the estimation process produced stable estimates.

46. The R&D Survey and the Energy Use Survey produced early estimates from about 60 percent response with additional iterations. Again the only unusual estimates were explained by very high weights, which in a final run would have been avoided by merging weighting cells or by special treatment. They were also limited by not being able to run the production version of the imputation due to the system configuration. Because of this estimation was done largely manually. However, the process proved useful for analysts in understanding the details of the estimation phase rather than just running the estimation process once on a final dataset.

47. The ICT Supply Survey did one run of the early estimation but they were prevented from doing more iterations due to having to spend their time implementing the new weighting methodology.

48. The Screen Industry Survey had the best results from early estimation, at about 55 percent response. The results were good because the estimates highlighted several errors where the decimals corresponding to cents on the questionnaire were being read as extra digits, meaning some responses were 100 times larger than they should have been. While these errors would have been identified eventually, recognising them early meant that deadlines were not put under pressure later in the process.

49. *Top-down editing.* Overall, while some surveys did some form of top-down editing, the extent that this approach was used could be improved by making estimation processes more flexible. This flexibility would allow analysts to easily produce estimates and sample errors for subpopulations for which final estimates may not be produced.

50. The Business Operations Survey drilled down from estimates at industry level to sub-industry level and then to a unit-record level. Similarly, the Screen Industry Survey produced estimates at industry level then for the contractor subgroup and then reviewed unit records. Neither of these surveys considered estimates by stratum or weighting cell.

51. The Energy Use Survey did look at ranked lists by stratum after drilling down from suspicious industry estimates. However, the process could have been improved by first producing estimates by strata to identify the most suspicious strata to review first.

52. *Use of sampling errors to identify suspicious estimates.* This approach worked well for the Business Operations Survey; it could have worked better for the Energy Use Survey but could not be done for the R&D Survey.

53. The Business Operations Survey checked sample error estimates against those from previous years which did not reveal any problems. The Energy Use Survey produced sample error estimates only as part of the final estimation process, which was not early enough to be used to identify any data errors. However, it did highlight a potential improvement to the sample design process for future years. The R&D Survey did not have their sample error estimation code working early enough to produce estimates for macro checking.

54. The Screen Industry Survey is full coverage and although weight adjustment is used to deal with non-response, sample errors are not usually estimated. However, rough estimates were created this year, using the surveymeans procedure is SAS. The purpose of these ad hoc sample error estimates were to give an indication of whether the response rate was high enough to close off, or to continue pushing for responses past the original close-off date. Not only did these ad hoc sample errors answer the response rate question, but they also helped identify more cases of the decimal place error as mentioned above.

55. *Top contributor method.* Ranked lists were already used by most business performance surveys; however, it is possible that in some cases, there is no clear understanding of why they were being used.

56. All five surveys produced lists of top contributors and biggest changes. These lists ranged from top five to top 30 and were for a range of key variables.

57. The Screen Industry Survey and the Business Operations Survey also compared the top contributors lists with those from previous years. This comparison was made to understand the data and what had changed. However, as a result of these comparisons the Business Operations Survey called back some unusual respondents that had, for example, gone from a response of $10 million to zero.

58. The ICT Supply Survey determined the size of their lists by considering how much the units on the list contributed to totals. For ICT sales, the top 30 contributed about 60 percent of the total estimate. For other estimates of sales by commodity, it was possible to draw the line at as low as five contributors and still have the units on the list contributing up to 80 percent of the total.

59. The Energy Use Survey targets different industries each year and, of the industries in 2010, only the manufacturing industry had been surveyed previously. As such it was only possible to produce a list of the largest changes for the manufacturing industry. However, even this proved to be a challenge as the previous data was collected for a different level of reporting unit. Looking at top contributor lists for the industries that hadn't been surveyed previously not only identified some data errors but was also a key way of gaining more subject matter knowledge about energy use in these industries.

60. *Graphical analysis.* Interestingly, graphical analysis was not heavily used by any of the surveys as the analysts found that they were able to sufficiently interpret the data by looking at lists, unit record files, or tables. The Energy Use Survey did attempt to produce some scatter plots; however, scale issues meant it was too difficult to interpret anything about unit records.

61. *Processing checks.* In the past, processing checks have generally always been done by the business performance surveys. The change in practice was more about trying to capture them in one place and ensure that they are feeding into macro checking. None of the surveys mentioned any problem with performing the processing checks and none identified any system problems through the processing checks. It was mentioned that automation of the creation of all or some of the processing checks would be useful.

C. Quality indicators

62. As mentioned previously, it is not currently possible to calculate all of the indicators for all five business performance surveys due to difficulties in compiling the data required. In some cases, only estimates of the actual value were possible. Even the indicators that are not estimates are subject to some assumptions in the calculation process (such as assuming clerical edits have been accurately noted). No indicators have been calculated or estimated for the Screen Industry Survey.

Table 4

Quality indicators for the business performance surveys

Quality indicator	Business Operations Survey		R&D Survey		ICT Supply		Energy Use	
	2009	2010	2008	2010	2008	2010	2009	2010
Edit failure rate (overall)	0.96	0.77	1.30	1.58	2.16*	2.38*	0.39	0.23
Average number of edit failures per unit			5.72	6.66	5.00*	4.23*	1.34	1.00
Clerical hit rate	1	1	1	1	1	1	0.61	0.6*
Clerical edits	4130	2497	2619	787	4000*	900*	1206	740
Respondent recontacts	20*	20*			20	12	125*	50*
Error rate by error cause								
Duration of E&I (work days) – Micro	104	83	56*	68*	50*	20*	50*	77*
Duration of E&I (work days) – Macro	2	10			15*	20*	15*	80*
Resource usage (staff hours) – Micro	337.5	187.5			375*	150*	375*	100*

Resource usage (staff hours) - Macro	15	40.5			112.5*	150*	188**	375**
* Estimate ** Estimate which also includes resource usage on imputation.								

63. The edit failure rate increased for two surveys and decreased for two surveys. Likewise, the average number of edit failures per unit decrease for two of the surveys and increased for one (and could not be measured for one). The differences from year to year for each survey are due to factors such as an increase in the number of edit checks or changes in the questionnaire. These two indicators are not particularly useful for reviewing the new approach on their own but should be considered in conjunction with other indicators.

64. The clerical hit rate is one (or 100 percent) for all surveys other than the Energy Use Survey. This shows that all of the alerted errors result in a manual adjustment. The Energy Use Survey has a value of less than one since some edit rules alert records that may be unusual but are often confirmed as valid once inspected.

65. The most interesting quality measure is the number of clerical edits. This dropped significantly for all surveys even though two of the surveys had an increase in the edit failure rate. This shows the large increase in the number of edit failures that were automatically edited.

66. The duration of E&I indicators show that for two surveys the micro editing actually took longer in terms of elapsed time despite the reduction in the number of clerical edits. However, this does not capture intensity of work during this time. For example, the Energy Use Survey may have had two analysts working full time on reviewing alerted micro edits in 2009 but only one analyst in 2010. The elapsed time also includes time taken to set up the new edit rules.

67. A better indicator for reviewing the new editing approach is the resource usage. All measured surveys show a large decrease in time spent (in terms of resource hours) on micro editing and a large increase in time spent macro editing. This is as intended and is consistent with the anecdotal review of the new approach.

VIII. Discussion

68. The implementation of the new approach to editing has proved more efficient and has provided the analysts with the direction and big-picture perspective which was the expected outcome of the quality reviews that first proposed the new approach. The new approach has seen the automation of many micro edits that were previously dealt with manually. It has also seen an increased amount of time spent macro editing as well as a wider range of macro editing approaches used.

69. The review of the implementation of the new approach showed that some of the recommendations have been successful while other aspects of the new approach could be improved.

70. The use of more automatic edits has been the main source of efficiency gains. Generally, they have led to large reductions in time spent manually micro editing. This can be further reduced in future years given the ability to reuse the automatic edit rules that were developed in 2010 as well as potential refinements such as adjusting thresholds and questionnaire improvements.

71. The macro editing techniques that proved the most useful were calculating estimates during the data collection process and the use of ranked lists. The use of processing checks was previously done well but was made more efficient through the production of a survey-monitoring template.

72. Macro editing can be still be improved by refining the estimation process. The implementation of estimation programmes that can easily be run on partially validated datasets (including unit non-response adjustment) will make it much easier to produce the early estimates that proved useful where they were able to be calculated. These programmes should include the ability to create estimates that may not be included in final outputs to allow for top-down editing. They should also include some form of sample error estimation.

73. While the use of graphical analysis was limited, it is still seen as a useful complement to the other macro editing techniques. This could be improved by more guidance around the production and interpretation of graphs.