

**UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE**

CONFERENCE OF EUROPEAN STATISTICIANS

Work Session on Statistical Data Editing
(Vienna, Austria, 21-23 April 2008)

Topic (v): Editing based on results (post-editing)

NEW TOOLS FOR STATISTICAL DATA EDITING

Supporting Paper

Prepared by Anders Norberg and Gunnar Arvidson, Statistics Sweden

I. INTRODUCTION

1. The goal of selective editing is to minimize the amount of manual data review while maintaining an acceptable quality of the statistical output. We imagine the bias that is still in the estimates due to the fact that not all data has been intensively verified¹. If this so called pseudo-bias can be estimated, the acceptable editing process should generate a bias less than 20% relative to other survey errors, as a rule of thumb. This level of relative pseudo-bias can be allowed to vary for different parts of the output.

2. Selective editing can be used as a complement to “regular” editing to reduce the list of flagged data that are identified by suspicion only. Selective editing can also consider suspicion and potential effects simultaneously in an integrated procedure. We propose quite simple continuous propensity measures for suspicion, which more effectively can be used in a score function than a dichotomous indicator.

3. One erroneous input data value can have impacts on several output statistical values. This is so when output is spread by more than one classification variable, for example when wages are computed and presented by gender, education, occupation and industrial sector. Here it is necessary that the producer of statistics can assess the importance of different output tables and estimate the consequences of lack of quality on the statistics, from the users point of view.

4. At Statistics Sweden the development is in full progress to build general IT- tool for selective editing using the score function. This paper is a short version of the demands the statisticians have formulated for a box of tools in a draft paper “General editing method at Statistics Sweden 2008”.

II. CONFIDENCE IN THE NATIONAL STATISTICAL BUREAU

A. Respondents and personnel

5. It can have a negative effect on respondents and personnel if too many erroneous data values pass through without any reaction. A reputation that *Statistics Sweden will accept whatever data is supplied* hardly promotes the will to supply high quality data. In order for confidence in Statistics Sweden to be maintained, it is desirable to identify and recontact respondents who have supplied data that are strongly suspected to be erroneous. This quality conflict to some degree with the necessity to maintain low costs for production.

¹ In this line of thought we do not consider the potential introduction of errors in the verification process.

6. One indication of erroneous data is if exactly the same information is supplied on several survey occasions, though this may not be reasonable. Edits for such errors are difficult to construct. The demand on the number of surveys rounds with repeated values varies – and an issue is how to explore cold-deck data to find out. This is about finding inliers and the general concept of selective editing, is not suitable.

7. Partial nonresponse can have its roots in bad measurement tools but also in an easy way to keep away from respondents duty. Either we make a recontact for all respondents with nonresponse even if this is not justified from a short-term resource perspective or we treat them in selective editing. If, in short term statistics, it is a repeated procedure by a respondent not to deliver information/data, the impact on statistics should be estimated not only for one survey occasion but for a series for which the respondent is supposed to be in the sample.

B. Customers

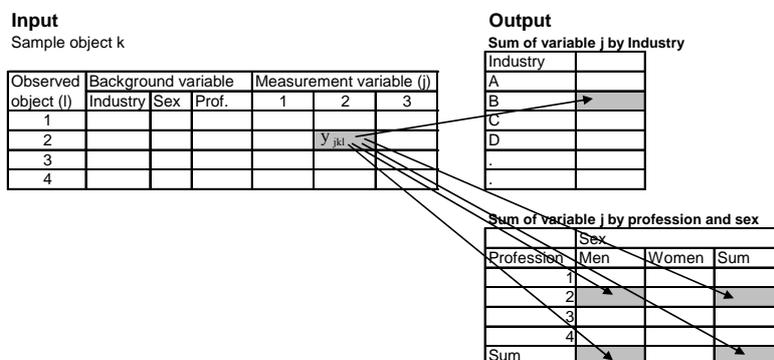
8. A customer that orders unplanned output of statistics or make use of micro-data might find suspected data that Statistics Sweden, focusing on costs and other presentations, has not verified. To foresee all possible uses of micro-data and by that see the impact of possible errors on all types of analysis is impossible. The expenses of actions to reduce errors in these sources cannot be covered by appropriation. A reasonable action, however, is to declare the quality of the micro-data. The users should be made aware of the number of suspected objects and the number of which have been reviewed and not reviewed. Most completely, process data could be attached to each data value with information on level of suspicion and actions taken (review or not, imputation etc.). Fatal errors are always put to rights by manual review or by imputation.

III GENERAL SETTING FOR SELECTIVE DATA EDITING

A. Statistical output

9. The person in charge of the statistics should provide a description of what is most important, evaluated in relative numbers. This information should be used to adjust parameter settings etc. in order to obtain high quality where it is most needed. A reduction of the manual verification and recontacts liberates resources of which some are proposed to be used for reducing other sources of error in the statistical process. Optimally this leads not only to a cost reduction but also to an improved overall quality.

10. When we have several classifications in the grouping of the statistical output into domains of study, we need to be able to globalize the potential and expected impacts respectively for data values obtained on the output as a whole. Wage and salary structures in the private sector is a statistical product with several classifications, by gender, education, occupation and industrial sector. On the opposite, several short-period statistical products have inclusion in overall the gross domestic product (GDP) as their most important final use. Here there is no need to evaluate impacts on the statistics in terms of classifications.



11. A classification is one or a few variables and specifications of values. The classifications can be constructed by complex mixtures of these variables. Margins of a cross-classification are single-classifications. A total is a very simple classification. Let $c = 1 - C$ denote classifications. A classification produces one or many subpopulations, domains of study. For a domain of study we can compute and

present statistics for; many variables and many measures. Let $d = 1 - D$ be domain of study. A statistical output value, a cell, is addressed by the combination $d_{.j}$ to tell both domain of study and variable.

B Variables

12. With some exceptions, such as for Structural business statistics, a survey includes a certain number of variables, the same number for each respondent. Some surveys are restricted to a single, main variable, such as price in price index statistics. When we have several variables, we want to evaluate the potential error of each variable for one observation object. We may then already have measured effects that are summarized for several classifications in the reporting.

13. The variables may have quite different properties for sorting, e.g. deliveries in kronor or kilograms, or employees' hours worked or wage totals in kronor (SEK). Here, we should relate the impacts of potential errors to estimates of the totals or standard errors in, for example, preceding surveys. We introduce a coefficient for the relative importance of each variable. A further reason for introducing such a coefficient is that predicted values vary in how well they function for the variables. A variable with lower predictive accuracy should have a lower coefficient so that it does not dominate in the manual review.

14. Some variables included in the score function may be strongly correlated. This can be a matter of a derived variable that is the sum of its component parts. What effect this has on the result of prioritizing objects depends, inter alia, on how global scores are constructed. Summing of local scores may have greater importance than using maximum local scores. *Further analysis is needed into correlated variables.*

C Several observations (records) per respondent

15. As exemplified by the foreign trade statistics and price statistics, it is common for respondents to have a highly variable number of observations. Thus, editing leads to respondents obtaining a variable workload of manual review. It would be an attractive property of an editing system if it could focus on minimizing this total workload for respondents for a given effect on statistics. For practical reasons, though, it might be necessary to focus on primary sampling units.

IV BASIS FOR COMPUTING SUSPICION AND POTENTIAL IMPACT

A Notions

16. We aim at computing a global score for each respondent (sample object) based on products of suspicion for the received data values and the potential impacts on the output statistics. In traditional editing suspicion is zero or one, accept or failure (flag). We propose a generalization towards a continuous measure 0-1.

17. At survey occasion t let for primary sampling unit or respondent k ($k = 1, \dots, n$) and secondary / observed unit l ($l = 1, \dots, n_k$)

$y_{j,k,l}$ = unedited (raw) value for variable Y_j ($j = 1, \dots, J$)

$\tilde{y}_{j,k,l}$ = "expected" value for variable Y_j ($j = 1, \dots, J$), based on any model using data for previous survey occasions or anecdotic information

$y_{j,k,l}^{(e)}$ = edited value for variable Y_j ($j = 1, \dots, J$), for data that were not reviewed, $y_{j,k,l}^{(e)} = y_{j,k,l}$

$w_{k,l}$ = weight for observed unit l ($l = 1, \dots, n_k$) within primary sampling unit k

w_k = weight for primary sampling unit k ($k = 1, \dots, n$).

18. The sample weight w_k refers in the first instance to design weight and/or calibration weight. If calibration or other methods of compensating for non-responses are used, there is no knowledge of the final weights prior to table production. Then, the editing process must be guided by the design weight,

possibly inflated in the light of experience from earlier surveys. For various reasons the weights might be specific for different variables Y_j ($j = 1, \dots, J$), but this is not common. For means the weights contain the number of units in the denominator ($1/N$).

B Using previous data (cold deck data)

19. The elementary components of the global score need good estimates of expected values and dispersion. These estimates must be computed before the editing of survey round t , at least if you want to edit the first entered data record. Calculations are based on edited data, not using raising factor, for homogenous groups. All objects are included in the groups for cross-sectional data no matter if they belong to an old outgoing sampling panel or the present one.

20. A decision has to be made whether to include imputed data or if they must be excluded. It seems most advisable from a theoretical point of view not to use imputed data, but it is easier not to make a difference. A decision must also be made whether to include data that was suspicious in the previous survey but not reviewed because the potential impact was low, or not. Again it seems to be a good idea not to use highly suspected data, but it is easier not to make a difference

21. For time series data several survey rounds are needed. For cross-sectional data one survey round can do, but several can be preferable. The data should be as fresh as possible.

22. For one-time-surveys we can possibly wait until half of all the records have arrived and been entered. Use what is available, divided into homogeneous groups and compute measures of central tendency and dispersion as above. This can be done continuously as more data are stored.

C Constructing homogenous groups for cross-sectional measures

23. It is necessary to construct homogenous groups for which expected values and dispersion intervals are computed. It is our empirical view, so far, that it is more important to stress homogeneity rather than a large number of records in the calculation of expected values for the purpose of editing.

24. Homogenous groups may, but need not, correspond to strata or domains of study. The groups can in a generalized way be formed by a set of classificatory variables, the detail of classification (number of digits) and a parameter stating the minimum number of observations required for the computation. We introduce H classificatory variables in the cold-deck data. They are accompanied by H parameters stating the number of digits.

25. The homogeneity groups are constructed according to a hierarchical tree-structure where splits are defined by the H classificatory variables as long as there are at least a minimum number of cold-deck records in the groups. It is no problem to let the identification of enterprises be a homogeneity variable. It is no problem to have the same variable at several levels with different number of digits in for example product and industry classifications.

D Central and dispersion measures

26. There are different approaches to get expected values and relevant dispersion intervals:

1. Time series data for the observed object k,l
 - a) Making use of the time variable in the model for prediction
 - b) Do not use time in the model for prediction
2. Cross-sectional data for a homogenous group of objects similar to k,l .

27. We propose the use of time series and cross-sectional measures in different surveys, setting the time series measures at priority and the cross-sectional as reserve when there is not a long enough time series. A set of measures for expected values and dispersion measures is proposed. Medians and quartiles across the survey rounds and across observations k,l for homogeneous groups per group are suggested.

Prediction for observation k,l and variable j can be made on the basis of regression analysis with explanatory variables. In SAS there is also an easy-to use procedure for robust regression which can be advantageous in this case.

28. In most surveys it is not necessary to use complicated time series models. A simple average of the last three survey rounds can do. If the survey measures phenomena with a heavy seasonal pattern, such as to consider the production decline in June and vegetable prices, steps should be taken to pay regard to this. SAS has a procedure for time series analysis with default modelling of an arima-model, the UCM-procedure. Using these models we get a predicted value (the expected value) and also a confidence interval to be easily used.

V SUSPICION

A. Fatal edits

29. We presuppose that measures have been taken for non-valid values, including non-response, in classification variables. This can be done by recontacts or imputation. Fatal edits identify

1. inconsistencies in continues variables (by balance edits),
2. item non-response,
3. non-valid values.

30. All these three types of checks result in flag or no flag, i.e. suspicion =1 or 0. In some surveys there are variables where it is impossible to distinguish between partial nonresponse and the value zero. There is no useful method for editing and imputation to deal with this situation. This is a measurement issue and the questionnaire must be adjusted to reduce this problem. When it comes to editing and imputation you have to know if there is a partial nonresponse.

B. Traditional query edits

31. A **test variable** is an arithmetic expression, based on the survey's variables, the object's sample weight and data from earlier production rounds and register data. The value of the test variable is compared with an **acceptance region**. In order to efficiently be able to identify suspect data values, it is essential that acceptance limits take account of large differences between **groups**, such as various sectors of industry and groups of varying size. The edits can, as suggest above, be based on time series or cross-sectional analysis. The edits result in flag (suspicion=1) or no flag (suspicion=0).

C. A general measure of suspicion

32. Let $t_{j,k,l,r}$ = derived test variable that is a function of unedited values $y_{j,k,l}$ and most often other variables, for example register variables. We introduce, at maximum, R test variables for each variable j , $r = 1, \dots, R$.

$\tilde{t}_{j,k,l,r}$ = "expected" value for the r :th test variable based on any model and previous data or anecdotic information, before the verification

$\tilde{t}_{j,k,l,r}^L$ = lower limits for some measure of "confidence" of $\tilde{t}_{j,k,l,r}$

$\tilde{t}_{j,k,l,r}^U$ = upper limits for some measure of "confidence" of $\tilde{t}_{j,k,l,r}$

33. It may be the case that $t_{j,k,l,r} = y_{j,k,l}$ or $t_{j,k,l,r} = \log(y_{j,k,l})$ where $y_{j,k,l}$ is a price ratio in a price index survey. Often a test variable is the difference or ratio between a received value and a corresponding value from a register or the previous survey. In the foreign trade statistics, price is denominated in quantities (kg, items, etc.) to assess suspicion, whereas the impact on the statistics is measured solely in kronor (SEK). Suspicion is independent of domain of study and the weights w_k and $w_{k,l}$.

34. The ratio $\text{Rat}_{j,k,l,r}$ is proportional to our r :th suspicion for data item j,k,l :

$$\text{if } t_{j,k,l,r} < \tilde{t}_{j,k,l,r} - \text{KAU} \cdot (\tilde{t}_{j,k,l,r} - \tilde{t}_{j,k,l,r}^L) \\ \text{Rat}_{j,k,l,r} = (\tilde{t}_{j,k,l,r} - \text{KAU} \cdot (\tilde{t}_{j,k,l,r} - \tilde{t}_{j,k,l,r}^L) - t_{j,k,l,r}) / \max\{(\tilde{t}_{j,k,l,r}^U - \tilde{t}_{j,k,l,r}^L), \text{MJAU} \cdot \tilde{t}_{j,k,l,r}\} \quad (1.1)$$

$$\text{if } \tilde{t}_{j,k,l,r} - \text{KAU} \cdot (\tilde{t}_{j,k,l,r} - \tilde{t}_{j,k,l,r}^L) < t_{j,k,l,r} < \tilde{t}_{j,k,l,r} + \text{KAU} \cdot (\tilde{t}_{j,k,l,r}^L - \tilde{t}_{j,k,l,r}) \\ \text{Rat}_{j,k,l,r} = 0 \quad (1.2)$$

$$\text{if } t_{j,k,l,r} > \tilde{t}_{j,k,l,r} + \text{KAU} \cdot (\tilde{t}_{j,k,l,r}^U - \tilde{t}_{j,k,l,r}) \\ \text{Rat}_{j,k,l,r} = (t_{j,k,l,r} - \tilde{t}_{j,k,l,r} + \text{KAU} \cdot (\tilde{t}_{j,k,l,r}^U - \tilde{t}_{j,k,l,r})) / \max\{(\tilde{t}_{j,k,l,r}^U - \tilde{t}_{j,k,l,r}^L), \text{MJAU} \cdot \tilde{t}_{j,k,l,r}\} \quad (1.3)$$

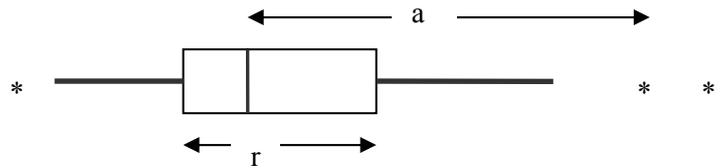
35. The parameter KAU is a means to decide how much of the sample that will have zero suspicion. More focus on potential impact, even when data looks reasonable, is achieved when $\text{KAU}=0$ and $\text{Rat}_{j,k,l,r} > 0$ for all data in practice. Focus on a minor part of data that is really suspected is achieved when $\text{KAU} > 2$. The parameter MJAU is necessary to make the ratio computable when the dispersion range is zero. By empirical experience it is suggested to be in the region of 0,05.

36. Through the transformation $\text{Susp}_{j,k,l,r} = \text{Rat}_{j,k,l,r} / (\text{TAU} + \text{Rat}_{j,k,l,r})$ (2) we obtain a measure of suspicion that always lies between 0 and 1, where the parameter $\text{TAU} > 0$.

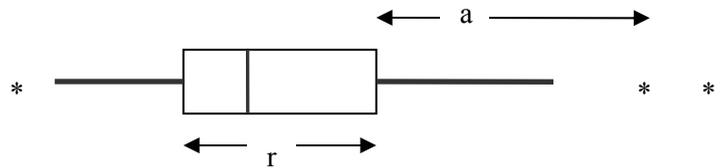
37. The last step, when there are $R > 1$ cross-sectional test variables, is to take the maximum over r of the $\text{Susp}_{j,k,l,r}$.

37. **Special case 1:** $\text{KAU}=0$. The ratio is the distance (a) from the centre $\tilde{t}_{j,k,l,r}$ divided by the dispersion range

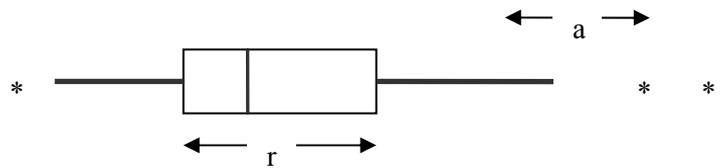
$$r = \tilde{t}_{j,k,l,r}^U - \tilde{t}_{j,k,l,r}^L, \quad \text{Rat}_{j,k,l} = a/r$$



38. **Special case 2:** $\text{KAU}=1$. The ratio is the distance from the nearest range limit divided by the range. For data between the lower and upper limits of the dispersion range the suspicion is 0.



39. **Special case 3:** $\text{KAU}=2$. For data between the centre minus two times the distance between the centre and the lower limit and the centre plus two times the distance between the upper limit minus the centre the suspicion is 0.



40. **Special case 4:** $\text{KAU}=1$ and let TAU be a very small number, for example $\text{TAU}=0,001$. The ratio $\text{Rat}_{j,k,l,r}$ is zero when the test variable value is between the dispersion limits and $\text{Susp}_{j,k,l,r}$ is zero. The ratio $\text{Rat}_{j,k,l,r}$ is positive when the test variable is outside the interval and $\text{Susp}_{j,k,l,r}$ is close to one, regardless of $\text{Rat}_{j,k,l,r}$.

D. Overall suspicion

41. Suspicion for a data value j,k,l is set in a four-step order:

- If fatal edits are in place and at least one of them flag for error, suspicion for the data item is 1.
- If a) did not result in suspicion = 1 and at least one traditional query edit is specified, and at least one of them flag for error, suspicion for the data item is 1.

- c) If neither a) nor b) result in suspicion = 1 and a time series edit is specified, the result of this is the suspicion.
- d) If neither a) nor b) result in suspicion = 1 and no time series edit is specified and one or several general query edits are performed, we propose taking the maximum suspicion of them.

VI IMPACT

A. General

42. For each cell d,j we can estimate from a census or a sample the totals $\hat{Y}_{d,j}(t)$ for the reference period t . In most surveys this estimate is a linear combination of observed and edited data $y_{j,k,l}^{(e)}$ from period t . If so, the actual impact on the estimated sum for any cell d,j is the weighted difference between the raw data item and the edited item $w_k \cdot w_{k,l} \cdot (y_{j,k,l} - y_{j,k,l}^{(e)})$. In case of non-linear estimators, such as ratios, a Taylor's expansion to a linear expression for impact often suffices. We then have $w_k \cdot w_{k,l} \cdot (y_{j1,k,l} - y_{j1,k,l}^{(e)} - r \cdot (y_{j2,k,l} - y_{j2,k,l}^{(e)}))$ where r is the ratio of totals for variables Y_{j1} and Y_{j2} .

43. The actual impact is unknown until verification. We therefore introduce potential impact, where the edited data $y_{j,k,l}^{(e)}$ is replaced by the expected data $\tilde{y}_{j,k,l}$. The absolute value of potential impact of $y_{j,k,l}$ before editing when it is included in a domain d is:

$$\text{Pot Im}_{d,j,k,l} = w_i \cdot w_{i,k} \cdot |y_{j,k,l} - \tilde{y}_{j,k,l}| \quad (3.1)$$

and analogously for ratios.

44. The expected impact is suspicion multiplied with potential effect:

$$\text{Exp Im}_{d,j,k,l} = w_k \cdot w_{k,l} \cdot (y_{j,k,l} - \tilde{y}_{j,k,l}) \cdot \text{Susp}_{j,k,l} \quad (3.2)$$

and analogously for ratios.

45. Let $\hat{Y}_{d,j}(t-1) = \sum_{k,l} w_k \cdot w_{k,l} \cdot y_{j,k,l}^{(e)}(t-1)$ be an estimate of the total for variable Y_j on reporting for the domain of study d,j where observation unit k,l is included, from a previous survey² and $\hat{\sigma}_{d,j}(t-1)$ = an estimate of the standard error of $\hat{Y}_{d,j}(t-1)$. The relative expected impact is (3.2) divided by estimated total or its standard error.

B. Impact from fatal errors

46. The score function is primarily used for statistical edits, but it is also possible to include non-statistical edits, but this is not straight forward. Flags from fatal edits always contribute to the suspicion measure ($\text{Susp}=1$). In case of recontact, fatal errors should always be corrected.

47. An edit which checks that a total equals the sum of its parts is a balance edit or accounting edit. Flagged data from balance edits are corrected in the manual review if the global score exceeds the threshold value, otherwise by automated imputation. If the sum of the variables a , b and c should equal d , then the balance edit will be based on the test variable $t = d - a - b - c$. Local scores are calculated for the test variable t where the impact is related to the variable d and the suspicion is set to 1. Local scores are also normally calculated for each of the variables in the balance edit and are included in the global score.

48. Item non-response has to be handled by follow-up and/or by imputation, manually or automatically. Class variables with non-response must always be imputed (provisionally or finally)

² May be the corresponding month one year earlier or the mean over several months.

before scores are computed. This is necessary as the score is related to the specific cell the data value belongs to and the score cannot be computed otherwise.

49. For continuous variables there are two alternative approaches that can be used, which can be seen as different ways to take impacts on estimates into consideration. One is to set up conditions where influential objects go to manual review or imputation, irrespective if the global score for the object is under or over the threshold for other reasons. The other is to include this fatal error in the score function (Susp=1). A pre-specified, but arbitrary, potential impact is used for the item non-response in the score function. If the global score exceeds the threshold value the unit goes to follow-up, otherwise to automated imputation of the non-response.

50. To assess the gravity of the missing $y_{j,k,l}$ we can replace it with $\tilde{y}_{j,k,l}^L$ which is the lower limit for some measure of “confidence” of $\tilde{y}_{j,k,l}$. Hence, the potential impact from (2a) would then be replaced by $w_k \cdot w_{k,l} \cdot \left| \tilde{y}_{j,k,l}^L - \tilde{y}_{j,k,l} \right|$. However, the user may in some cases prefer not to let item non-response for a certain variable y_j contribute to the score function. To account for this situation we introduce a parameter VOF_j which can be set by the user. The whole relation (2a) is then replaced by

$$\text{Exp Im}_{d,j,k,l} = w_k \cdot w_{k,l} \cdot \left| \tilde{y}_{j,k,l}^L - \tilde{y}_{j,k,l} \right| \cdot 1 \cdot VOF_j. \quad (3.3)$$

51. The default value for VOF_j equals 1 while $Susp_{j,k,l}$ always equals 1 for fatal errors. When a value $y_{j,k,l}$ exists but is considered as a fatal error then (2a) amounts to

$$\text{Exp Im}_{d,j,k,l} = w_k \cdot w_{k,l} \cdot \left(y_{j,k,l} - \tilde{y}_{j,k,l} \right) \cdot 1 \cdot 1, \text{ where } Susp_{j,k,l} \text{ and } VOF_j \text{ have both implicitly been set equal to 1.}$$

C. Invalid entries in classification variables

52. Data errors are not necessarily found only in measurement variables. In many cases class variables are hard for the respondent to code. Examples include occupational coding in the Wage and salary structures in private sector and goods coding in foreign trade statistics. In case of invalid entries it is necessary to impute valid codes before calculations in the score function to be able to analyse the impact on estimates. Also item nonresponse will be imputed for the class variables. As the imputed value might be wrong the record is suspicious when calculating the scores.

53. In the case where a class variable will be used in a register with a variety of output, the product manager might prefer manual follow-up for all invalid entries in class variables. On the other hand this argument is not reasonable in the case when there are a large number of units involved.

54. During the manual review the editors also have to take into account that a deviating measurement value could be acceptable in the case where the class code is incorrect as for example an extreme wage for an occupational group or an extreme price per kilo for a traded good. We do not know in advance whether it is the measurement variable or the classification variable that is most likely to be at fault.

VII SCORE FUNCTION

A. Importance parameters

55. The design of the selective editing can be adjusted in order to capture survey-specific needs. To set these edit-parameters judgmentally good knowledge about the survey is needed. Thus it is recommended to involve the person(s) in charge of the survey at this stage. The default value for all of these is set to 1, which will do for many surveys.

56. $VIOLIN_j$ = relative importance of variable Y_j ($j = 1, \dots, J$).

$CLARINET_{c(d)}$ = relative importance of the classification c ($c = 1, \dots, C$) in reporting of the statistics for the domain of study d .

57. $OBOE_j$ = importance of the size of the domain of study, $0 \leq OBOE_j \leq 1$, ($j = 1, \dots, J$). $OBOE_j$ is used to compute either a potential impact in absolute terms or a relative an impact relative to size of the domain of study or, thirdly, anything in between of these two pure alternatives. For most statistical surveys $OBOE_j$ can preferably be less than 1, say 0,5.

58. $CELLO_{d,j}$ is an object-independent parameter, constructed of the three above, specified for each output cell d,j . It transforms the expected impacts for variables on different scales and variation to comparable levels. Besides the importance parameters $CELLO_{d,j}$ contains the estimated total from a previous period, $\hat{Y}_{d,j,t-1}$, its corresponding estimated standard error, $\hat{\sigma}_{d,j,t-1}$, and parameters $ALFA_j$.

$$CELLO_{d,j} = \frac{VIOLIN_j \cdot CLARINET_{c(d)}}{\left(\text{maximum}\{ALFA_j \cdot \hat{Y}_{d,j,t-1}, \hat{\sigma}_{d,j,t-1}\}\right)^{OBOE_j}} \quad (4)$$

59. $ALFA_j$ is the way to register what quantity the expected impact is related to, this is a choice made by the product manager. Because of the risk that some domain of study can have the estimated standard error = 0, it is wise to at least set a small $ALFA_j$, for example $ALFA_j = 0,05$. Setting $OBOE_j = 0$ means that only the nominator will contribute to the value of $CELLO_{d,j}$. The contribution of the nominator will decrease with increasing values of $OBOE_j$. Setting a large $ALFA_j$ means that the bias is relative to the sum of Y_j .

B. From local to global scores

60. The score function can be divided into three parts; the suspicion (S), the potential impact (PE) and $CELLO_{d,j}$. Notice that the $CELLO_{d,j}$ -parameters can be regarded as weights applied to the potential impacts in order to make them comparable. The score on the most detailed level is

$$L1Score_{d,j,k,l} = |\text{Exp Im}_{d,j,k,l}| \times CELLO_{d,j} \quad (5.1)$$

61. The scores are globalized step by step, over domains of study (d), variables (j) and observations (l) to respondent or primary sampling unit (k) as:

$$L2Score_{j,k,l} = \left(\sum_d \left(\max\{0, L1Score_{d,j,k,l} - T_{L1}\} \right)^{GAMMA1} \right)^{1/GAMMA1} \quad (5.2)$$

$$L3Score_{k,l} = \left(\sum_j \left(\max\{0, L2Score_{j,k,l} - T_{L2}\} \right)^{GAMMA2} \right)^{1/GAMMA2} \quad (5.3)$$

$$GScore_k = \left(\sum_l \left(\max\{0, L3Score_{k,l} - T_{L3}\} \right)^{GAMMA3} \right)^{1/GAMMA3} \quad (5.4)$$

where T_{L1} , T_{L2} and T_{L3} are local threshold values and $GAMMA1 - GAMMA3$ are parameters, see Hedlin (2008).

62. We recommend as default value $T_{L1}=0$. The default local threshold values T_{L2} and T_{L3} can be tenth of the global threshold value.

Special case 1: Let local threshold = 0 and $GAMMA1=1$, we get the pure sum function

Special case 2: Let local threshold > 0 and $GAMMA1=1$, we get sum of local scores with a deduction for a certain threshold value

Special case 3: Let local threshold = 0 and $GAMMA1=2$, we get the sum of squares function

Special case 4: Let local threshold = 0 and $GAMMA1=10$, we get a function very close to maximum

VIII SETTING THE GLOBAL SCORE

63. A tool for selective editing according the specifications in this paper has many parameters to be set. Several of these can be set to the default values. We see today no other option than empirical studies within this concept to reach best performance, while fulfilling the demand on maximum 20 percent relative pseudobias or a similar demand.

64. The method can also be applied more flexibly by ensuring that editing for each statistical production round in some practical manner, prioritises objects with high scores, and is concluded when resources or time are exhausted, or when it is not deemed profitable to continue the manual work.

A. New respondents

65. In short-term statistics the samples are rotated slowly, which is necessary for the estimation of change. It is desirable in editing to quickly be able to resolve any problems for new respondents so as on how the data should be supplied. We might wish to have a smaller acceptance region for the new respondents in the samples. There is a lack of access to equally satisfactory predicted values for the new as for respondents that have been involved previously. This means that the computed potential impact on the statistical output will be relatively high, and that more recontacts will be made. The measure of suspicion can, however, work against this, so that a score that considers both suspicion and potential impact will be of the same magnitude as for other objects. Special selections of new respondents can therefore be motivated.

B. Probability sampling of suspected objects

66. We have no information on errors in the data that have not been reviewed, for respondents with a score under the global threshold. Hopefully, selective editing is efficient enough to decrease the total number of reviewed objects, compared to traditional editing, but this is a disadvantage when it comes to evaluation. One possibility is to intensively review a larger body of material from time to time. As a basis for adjusting the threshold value prior to a new survey round, verification should also be performed on items lying below the threshold value. This can be sample-based on intermittent basis.

67. In a survey with one variable (or just a few) and with one domain of study (or just a few) it should be possible to: (1) take an pps sample of the sampled objects as a sub-sample, (2) recontact the respondents in the sub-sample to review the correctness of suspicious data supplied, and (3) estimate the error in the point estimate with unedited data, so as (4) to perform a design-based adjustment of this estimate on the basis of the sub-sample. This long-winded process can be motivated if we know that errors tend to be not symmetrical, for example if it is understandable that respondent can forget some accounts in the total sum.

68. It is appropriate to select a sample of objects under the global score with probabilities proportional to the global score. In this application, it is reasonable to see to it that a large proportion of the objects have a positive score; in any case, there will be a few objects with low scores in the sample. A feasible method could be to make a Poisson sample without replacement with probabilities $p = n * \text{Score} / \sum \text{Scores}(t-1)$, where n is the targeted sample size, Score is the global score for the respondent and $\sum \text{Scores}(t-1)$ is the sum of scores under the threshold for some previous survey. Cut-off for the smallest scores can still be relevant. Poisson sampling is a suitable method as it can be used for the first to the last respondent regardless of inflow of data.

IX REFERENCES

- Adolfsson, C. and Gidlund, P. (2008), "Conducted case studies at Statistics Sweden", paper to be presented at the Editing workshop I Vienna April 2008
- Hedlin, D. (2008), "Local and global score functions in selective editing" paper to be presented at the Editing workshop I Vienna April 2008
- Latouche, M. and Berthelot, J.-M. (1992) "Use of a score function to prioritize and limit recontacts in business surveys", *Journal of Official Statistics*, Vol 8, No. 3 1992, pp 389-400
- Lawrence and McDavitt (1994) "Significance Editing in the Australian Survey of Average Weekly Earnings", *JOS* No. 4 pp. 437-447
- SCB (2002) "Guide till granskning", handbook in the series Current Best Methods from Statistics Sweden, CBM 2002:1