WP. 13
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

A MACRO SIGNIFICANCE EDITING FRAMEWORK TO DETECT AND PRIORITISE ANOMALOUS ESTIMATES

Invited Paper

Prepared by Keith Farwell and Paul Schubert, Australian Bureau of Statistics, Australia[1]

I. Introduction

1. This paper describes a framework for macro editing which uses scores based on significance criteria to detect anomalous estimates. It is an extension of the micro significance editing framework (a type of selective editing) used within the Australian Bureau of Statistics.

2. Statistical data editing may be defined as the activity aimed at detecting, resolving and treating anomalies in data to help make the data 'fit for purpose' (ABS, 2007). Whereas micro editing involves the editing of collection inputs such as unit records (or micro data), macro editing involves the editing of collection outputs such as estimates, ratios of estimates, and standard errors (i.e. macro data). For simplicity, this paper will refer to all collection outputs as 'estimates' unless a distinction is necessary.

3. The ABS is looking to introduce more objectivity into the macro editing process. One area of interest is the use of scores to detect and prioritise anomalous estimates. A score-based anomalous estimate detection process will add rigour and repeatability to the overall detection process (which may also contain a subjective detection element). The ABS has built a tool, called the Significance Editing Engine (SEE), which can be used to perform micro significance editing (Farwell, 2004; Farwell, 2005; ABS, 2011) on business surveys. This paper looks to extend the significance editing approach to the detection phase of macro editing. A measure of significance is used to develop a *macro significance score* which can be used to detect anomalous estimates and prioritise them.

4. A particular feature of the *macro significance editing* framework is that it provides information to allow the user to target macro editing effort where there is a hierarchy of estimates of interest involved e.g. national, regional, and sub-regional estimates. An application of *hierarchical* macro editing to survey data is provided.

[1] Views expressed in this paper are those of the authors, and do not necessarily represent those of the Australian Bureau of Statistics. Where quoted, they should be attributed clearly to the authors.

II. Macro editing scores and significance

A. Using scores in editing

5. Many basic scores that have been used in macro editing have the following general form

$$Score = \frac{Observed\ estimate - Expected\ estimate}{Scaling\ factor} \quad (1)$$

Various types of scores can be derived from (1) by substituting different choices for the expected estimate and the scaling value.

6. Note that the term 'estimate' may also refer to a rate (calculated as a ratio of two estimates of total) or an estimate of standard error (or coefficient of variation for a census). If the estimates used are rates, we obtain many typical ratio scores (which are a popular choice for macro editing scores). This paper will refer to the varieties of scores as:

(i) *estimate scores* (for estimates of total);
(ii) *ratio scores* (for ratios of two estimates of total);
(iii) *standard error scores* (for estimates of standard error); or
(iv) *coefficient of variation scores* (for coefficients of variation in censuses).

7. There are two major aspects that fundamentally affect the quality of scores based on (1) which are:
(i) the quality of the expected estimates; and
(ii) the choice of scaling value.
Importantly, the scores outlined above are not set up within a significance context. They take no account of the importance of the estimate being scored with respect to the totality of estimates requiring macro editing.

B. Predicted impact of editing

8. Significance editing is based on predicting the impact of editing actions on the outcomes considered important. Significance scores have the following general form

$$Score = 100 \times \frac{Measure\ of\ predicted\ impact\ of\ editing}{Scaling\ value} \quad (2)$$

9. In micro significance editing, the impact is measured as the bias in a set of chosen estimates caused by possible reported data errors (Farwell, Poole and Carlton, 2002). The basic measure of predicted impact for micro significance editing is:

Micro editing impact = Adjusted expected target estimate – Expected target estimate (3)

where the *adjusted expected target estimate* is calculated in the following way. The expected target estimate is a function of the expected unit values and the estimation methodology. When a reported value is obtained, we remove the contribution of its associated expected value from the calculation of the expected target estimate and replace it with the contribution from the reported value. That is, we replace the expected unit value with the reported unit value and we recalculate a new expected target estimate. This is done on a value by value basis. Accordingly, there is an adjusted expected target estimate value for each reported value requiring a score. For Horvitz-Thompson estimates of total, this is equivalent to multiplying the difference between the reported and expected value by the unit weight. Although the Horvitz-Thompson impact is obvious, (3) is needed to support more complex scores involving estimates such as rates, standard errors, indexes, etc.

10. The scaling value in (2) may be either an expected estimate or an expected standard error. The score can be considered to 'target' a set of estimates. These are referred to as *target estimates* while the domain containing them is referred to as the *level of significance* or *target domain* in this paper.

C. The concept of significance in macro editing

11. The definition and measurement of predicted impact for macro editing is far less straightforward than for micro editing. There are often conflicting macro editing priorities which change as macro editing progresses. For example, consider a situation where we have previous and current estimates for several data items at regional level. Our first priority is to produce high quality national level estimates (and high quality movement estimates as a by-product). However, we also want to produce good quality regional level estimates. How can we address both sets of priorities?

12. If we were not in a position to edit every regional estimate with a large movement, it would be logical to assess the importance of the regional movements in terms of their impact on the national movements. The following two scores which measure the regional movement for data item *i* as a percentage of the previous regional and national estimates, can be used to analyse the problem.

$$Region\_est\_score1 = 100 \times \frac{Y_{i,t,Region} - Y_{i,t-1,Region}}{Y_{i,t-1,Region}} \tag{4}$$

$$National\_region\_est\_score1 = 100 \times \frac{Y_{i,t,Region} - Y_{i,t-1,Region}}{Y_{i,t-1,National}} \tag{5}$$

13. There will be some regional estimates which have very large scores at regional level but very small scores at national level which indicates that their movements are relatively unimportant at national level. Other regional estimates have smaller regional scores but their national level scores indicate they are important at national level. The difficulty with using scores in macro editing is the need to balance conflicting priorities.Which is more important from a macro editing perspective?

14. There are often more than two levels of estimates to deal with. For example, in an ABS census of Australian agriculture, there may be up to 30 variables in any of 65 statistical divisions (subregion level), up to 300 variables in any of 8 States (regions), and over 900 variables Australia-wide (national). As the number of variables increase, and as the levels become finer, we find ourselves faced with a macro editing dilemma. It is not uncommon to be faced with thousands of estimates and movements across several levels which need to be assessed and prioritised during macro editing.

15. This paper proposes that the predicted measure of impact for macro editing be following extension of (3):

Macro editing impact = Adjusted expected target estimate – expected target estimate (6)

where the definition of the adjusted expected target estimate is the natural extension of the definition used for micro-editing. When a base estimate is scored, we remove the contribution of its associated expected estimate from the calculation of the expected target estimate and replace it with the contribution from the actual base estimate. That is, we replace the expected base estimate with the observed base estimate requiring a score and we recalculate a new expected target estimate. This is done on a base estimate by base estimate basis. Accordingly, there is an adjusted expected target estimate value for each base estimate requiring a score. For estimates of total, this is equivalent to using the difference between the observed and expected base estimate as the measure for macro-editing impact. Although the macro-editing impact is obvious for estimates of total, (6) is needed to derive more complex scores involving estimates such as rates, standard errors, and indexes.

16. A macro-editing score can be developed using (6). However, unlike the micro editing version, there may be several target levels and several scores associated with a single observed estimate. In fact, the predicted impact will depend on the level of significance and a scaling value for each estimate at that level will be required. This idea is developed in this paper by incorporating the concept of a

hierarchy of targets and *hierarchical* scaling values. The general form of a macro significance score is:

$$Score = 100 \times \frac{Measure\ of\ predicted\ macro\ editing\ impact}{Scaling\ value\ for\ target\ level} \quad (7)$$

and scores (4) and (5) are an example of a two-level hierarchy involving two estimate scores.

III. The macro significance editing framework

A. Outline

17. In this section, we outline macro significance concepts such as the domain of study, the level of significance, base and target estimates and scores, expected estimates, sensitivity measures, hierarchical scores, hierarchical macro edits, combined scores, cost/benefit curves and Gini indexes. Macro significance scores will be defined for estimates of total, ratios of estimates of total, and standard errors of estimates using (6) and (7). The framework allows scores to be combined (Farwell, 2010). For example, a current ratio score can be combined with estimate scores for the numerator and denominator estimates. It will be possible to rank estimates using functions of scores, functions of ranks when several ranks are involved, or functions of both scores and ranks. The use of expected estimates in the measure of predicted impact leads to more complex scores such as a current ratio score which uses historical estimates as expected values for the numerator and denominator estimates.

B. Base scores and the study domain

18. *Base* scores are scores where the scaling value in (7) is from the study domain. These scores require observed and expected estimates (or expected standard errors) at the base level. The expected estimates may be based on historical estimates, modelled estimates, current medians or averages, or as a last resort, guesses. *Region_est_score1* (4) is an example of a base score.

19. The base estimate score is:

$$S_{est,base}(Y_i) = 100 \times \frac{\Delta(Y_{i,base})}{Y^*_{i,base}} \quad (8)$$

where $Y_{i,base}$ and $Y^*_{i,base}$ are the observed and expected estimates of total for variable *i* within the base domain, and

$$\Delta(Y_{i,base}) = Y_{i,base} - Y^*_{i,base}$$

20. The base ratio score is:

$$S_{ratio,base}(R_{ij}) = 100 \times \frac{\Delta(R_{i,j,base})}{R^*_{ij,base}} \quad (9)$$

where:
$Y_{i,base}$ is the observed numerator base estimate of total for variable *i*;
$Y_{j,base}$ is the observed denominator base estimate of total for variable *j*;
$Y^*_{i,base}$ is the expected value for $Y_{i,base}$;
$Y^*_{j,base}$ is the expected value for $Y_{j,base}$; and

$$R_{i,j,base} = \frac{Y_{i,base}}{Y_{j,base}}$$

$$R^*_{i,j,base} = \frac{Y^*_{i,base}}{Y^*_{j,base}}$$

$$\Delta R_{i,j,base} = R_{i,j,base} - R^*_{i,j,base}$$

21. Base scores using expected estimates as scaling values cannot be defined when expected base

estimates are zero and expected standard errors should be used instead. If expected standard errors are used, replace the expected base estimates in the denominators of (8) and (9) with $\alpha_{base}\,\mathrm{SE}^*(Y_{i,base})$ and $\alpha_{base}\,\mathrm{SE}^*(R_{ij,base})$ respectively, where $\mathrm{SE}^*(Y_{i,base})$ and $\mathrm{SE}^*(R_{ij,base})$ are expected standard errors. The parameter α_{base} has been incorporated to allow the expected standard error to be converted to a *bias tolerance* (with $\alpha_{base} = 1$ suggested as the default value).

22. The base standard error score for an estimate of total is:

$$S_{se,base}(Y_i) = 100 \times \frac{\Delta SE(Y_{i,base})}{\alpha_{base} SE^*(Y_{i,base})} \tag{10}$$

where $$\Delta SE(Y_{i,base}) = SE(Y_{i,base}) - SE^*(Y_{i,base})$$

23. The base standard error score for an estimate of rate is:

$$S_{se,base}(R_{i,j}) = 100 \times \frac{\Delta SE(R_{i,j,base})}{\alpha_{base} SE^*(R_{i,j,base})} \tag{11}$$

where $$\Delta SE(R_{i,j,base}) = SE(R_{i,j,base}) - SE^*(R_{i,j,base})$$

and *i* and *j* represent two different variables.

24. Note that equivalent scores for censuses can be created based on observed and expected coefficients of variation.

25. The standard error score is interesting in that it is usually only those observed to be larger than the expected standard errors that are usually considered as anomalous. However, one could argue that a standard error that is much smaller than expected could also indicate a macro editing problem (such as a systematic processing error). If we are only interested in standard errors that are too large, we add the following conditions to (10) and (11):

$S_{se,base}(Y_i) = 0$ when $\mathrm{SE}(Y_{i,base}) < \mathrm{SE}^*(Y_{i,base})$ for standard error scores for estimates; and
$S_{se,base}(R_{ij}) = 0$ when $\mathrm{SE}(R_{ij,base}) < \mathrm{SE}^*(R_{ij,base})$ for standard error scores for rates.

26. If expected estimates are used as scaling values, replace the expected base standard errors in the denominators of (10) and (11) with $Y^*_{i,base}$ for estimate scores or $R^*_{ij,base}$ for ratio scores. Any standard error base score using expected estimates as scaling values cannot be defined when expected base estimates are zero and expected standard errors should be used.

27. Movement scores are not developed in this paper (though they could be a consideration for a collection designed specifically to measure accurate movements). They are not needed for most collections because movement scores are very similar to estimate scores which use previous estimates as expected estimates.

C. Sensitivity measures

28. From a technical viewpoint, the anomaly identification process can be subject to two kinds of identification errors referred to as *swamping* and *masking*. Swamping is said to occur when estimates which are not anomalies are declared as anomalies. Masking is said to occur if actual anomalies are not detected as anomalies. For further details refer to Gather and Becker (1997); Samprit, Hadi, and Price (1999); and Maimon and Rokach (2005).

29. To manage these problems affecting the quality of the scores, we introduce the concept of *sensitivity measures* which are an auxiliary layer of conditions imposed on the anomaly detection process.

They can be used to exclude specific estimates from the anomaly selection process or to modify the scores themselves. The conditions that Sigman (2005) placed on labelling initial anomalies as final anomalies in are an example of *conditional* sensitivity measures. The magnitude transformations used in the historical and current ratio Hidiroglou-Berthelot macro edits (Hidiroglou and Berthelot, 1986) are examples of *multiplicative* sensitivity measures. Multiplicative sensitivity measures tend to be implicit since they are generally included in the definition of the score. If we were to use base scores only for anomaly detection, some form of sensitivity measure would be needed to control size masking. Some examples of sensitivity measures are:

$$d_{i,base} = max\left(\left|Y_{i,base}\right|, \left|Y^*_{i,base}\right|\right)^U \text{ and } 0 < U \leq 1 \tag{12}$$

$$d_{i,base} = \left(\frac{max\left|Y_{i,base}\right|, \left|Y^*_{i,base}\right|}{\left|Y^*_{i,base}\right|}\right)^U \text{ and } U \geq 1 \tag{13}$$

where each could be combined with a base score (such as a standard error score) and anomalous estimates are those with $s_{i,base} > c_1$ and $d_{i,base} > c_2$ (c_1 and c_2 are cut-offs).

30. Relative sensitivity measures allow a single sensitivity cut-off to be applied to many variables. Other examples of conditional sensitivity measures include restrictions on the minimum number of respondents allowed for base estimates, restrictions on the minimum number of estimates contributing to target estimates, and the exclusion of estimates of zero.

D. The level of significance, target estimates, and hierarchical scores

31. Hierarchical scores are a specific multiplicative application of sensitivity measure (13) with U=1 to base scores where the end result is a score which uses a *target* estimate as the scaling value. That is, hierarchical scores are scores for base estimates where the level of significance is a higher level than the base level.

32. The hierarchical estimate score is:

$$S_{est,base,target}(Y_i) = 100 \times \frac{\Delta Y_{i,base}}{Y^*_{i,target}} \tag{14}$$

33. The hierarchical ratio score is:

$$S_{ratio,base,target}\left(R_{i,j}\right) = 100 \times \frac{R^*_{i,j,target|base} - R^*_{i,j,target}}{R^*_{i,j,target}} \tag{15}$$

where $$R^*_{i,j,target} = \frac{Y^*_{i,target}}{Y^*_{j,target}}$$

is the expected target ratio;

and $$R^*_{i,j,target|base} = \frac{Y^*_{i,target} + \Delta Y_{i,base}}{Y^*_{j,target} + \Delta Y_{j,base}}$$

is the adjusted expected target ratio (calculated along the lines suggested in paragraph 9.)

34. Hierarchical scores using expected estimates as scaling values cannot be defined when expected target estimates are zero. Difficulties can also arise with hierarchical scores when base estimates can be both positive and negative. For example, as the sum of expected base estimates approach zero the hierarchical score becomes increasingly erratic. It is recommended to use the standard error as the scaling value in such cases. If standard errors are used as scaling values, replace the expected target estimates in the denominators of (14) and (15) with $\alpha_{target}\,SE^*(Y_{i,target})$ and $\alpha_{target}\,SE^*(R_{i,j,target})$ respectively (using $\alpha_{target} = 1$ as the default).

35. Hierarchical scores for standard errors are affected by the independence of the base estimates. When they are not independent, target variances are not the sum of base variances leading to complicated scores.

36. *National-region-est-score1* (5) is an example of a hierarchical estimate score. It is common to have several levels of significance and, therefore, several hierarchical scores. Hierarchical scores are used to develop hierarchical macro edits which are described in the following section.

E. Hierarchical macro edits

37. *Hierarchical* macro edits can be used to detect anomalous base estimates while taking into account the importance of the base estimate deviations from their expectations in terms of their impacts on the chosen target levels. They involve a combination of base and hierarchical scores and cut-offs where a cut-off is chosen for each of the base and hierarchical scores. Although each cut-off can be chosen independently, the preferred method is to select the hierarchical cut-offs first and apply these to the base estimates. The distribution of those base estimates above the hierarchical cut-offs are then examined and a base cut-off is chosen. The hierarchical and base cut-offs are then applied to the full set of base estimates.

38. A value from (0,1) is assigned for each base estimate indicating whether it passed or failed each of the base and hierarchical edits (where '1' indicates the estimate 'failed' the chosen cut-off). Each base estimate is assigned an *n*-dimensional point where *n* is the number of hierarchies. For example, a three-level hierarchy results in points (or categories) such as (1,1,1), (1,1,0),….., through to (0,0,1) and (0,0,0). The first co-ordinate relates to the highest hierarchical level, the second co-ordinate, the next highest hierarchy, and so on. The last co-ordinate relates to the base level. For example, (1,0,1) indicates that the base estimate failed the highest level hierarchical edit, passed the second highest level hierarchical edit, and failed the base level edit.

39. The user can choose the hierarchical category or group of categories they feel is most appropriate to investigate. The anomalous estimates within each group can be ordered by the size of the base score or one of the hierarchical scores. For example, the category (1,1,1) would be top priority and typically ordered by base score size while those in (1,0,0) would tend to be ordered by the top level hierarchical score size. Hierarchical edits have the ability to address conflicting macro editing priorities while giving some flexibility to the macro editor. They can be combined with sensitivity measures if necessary.

40. Four types of prototype hierarchical macro edits are currently being tested in the ABS. One is for macro editing estimates of total and one for ratios. Both use an optional conditional sensitivity measure similar to (13) based on benchmarks. The third is also for ratios but combines the ratio and estimate score results using ellipsoidal distance. These require the user to provide expected base and target estimates. The fourth type is designed for use without expected estimates. It generates expected estimates through the use of a median. It uses an implicit multiplicative sensitivity measure similar to (13) based on benchmarks to create the hierarchical scores. Refer to Farwell (2009a) and Farwell (2009b) for more details.

41. These hierarchical macro edit prototypes revolve around six basic steps which are:

(a) create macro-data;
(b) create scores;
(c) create hierarchical cut-offs;
(d) apply hierarchical cut-offs and choose base cut-off;
(e) select hierarchical outlier categories; and
(f) select anomalous base estimates.

F Ranks and cut-offs

42. Various ranking methods are available within the *Significance Editing Manual* (ABS, 2011) and

this paper will not detail them. Cut-offs may be two-sided or one-sided. Two-sided cut-offs can be used when separate cut-offs are needed for each tail of the score distribution. This paper proposes that one-sided cut-offs be the default with an option for using two-sided cut-offs for combined scores.

IV. An example of the macro significance editing framework

A. An example of a three-level hierarchical macro edit

43. In this example, we demonstrate hierarchical macro editing on estimates of total from an ABS Agricultural collection using previous estimates for our expected estimates. It involves a three-level hierarchy consisting of statistical division (SD i.e. subregion) as the base level, State (region) as the next highest level, and Australia (national) as the highest level. The example dataset consists of 1646 SD estimates which aggregate to 290 State estimates and 49 Australian estimates involving 49 data items. We use (8) and (14) to calculate the following three estimate scores:

$$SD\_est\_score = 100 \times \frac{Current\ SD\ estimate - Previous\ SD\ estimate}{Previous\ SD\ estimate}$$

$$SD\_State\_est\_score = 100 \times \frac{Current\ SD\ estimate - Previous\ SD\ estimate}{Previous\ State\ estimate}$$

$$SD\_Aust\_est\_score = 100 \times \frac{Current\ SD\ estimate - Previous\ SD\ estimate}{Previous\ Aust\ estimate} \qquad (23)$$

44. In this example, SD (subregion level) is the base level, and we have two target levels that we are interested in, State (region level) and Australia (national level). The SD_est_score is the base score, SD_State_est_score is the base_target1 score and SD_Aust_est_score is the base_target2 score. In other words, there are three scores (each of the above) for each SD estimate. Because there are three scores, three cut-offs are required. Each SD estimate may fail one of the edits (i.e. have a score greater than the cutoff) and pass the other two, or have other combinations of passing and failing the three edits.

45. As outlined above, the SD-State and SD-Aust hierarchical cut-offs are the first to be chosen using graphs displaying score size verses rank (ordered in descending score size). Extreme scores are excluded from the graphs (to improve readability) by applying user-defined *graph cut-off* values (a default value of 100 is used). The excluded scores can be separately listed. For example, the ‘cost/benefit curve’ in figure 1 below was used to choose an SD_State_est_score cutoff of 1.75 (16 SD estimates were excluded from the graph). This resulted in about 1100 of the 1646 SD estimates being selected at this stage. Similarly, an SD_Aust_est_score cut-off of 0.25 was selected using a graph of SD_Aust_est_score size versus rank.

46. A similar process was followed to graph the cost/benefit curve for the distribution of *SD_est_score* – this was done twice, once prior to application of the two hierarchical cut-offs, and once after the two hierarchical cut-offs have been applied. The second of these plots was used to select 15.0 as the *SD_est_score* cut-off.

47. Using 0.25, 1.75, and 15.0 as the three cut-offs, Figure 2 below displays results for the hierarchical macro-edit categories. Macro-editors can choose the hierarchical categories they feel are most appropriate to investigate. The estimates within each group can be ordered by the magnitude of one of the three scores in (23). For example, category (1,1,1) should be top priority and the estimates within it should be ordered by descending *SD_est_score* size. Editors may wish to examine estimates within other categories with a view to augment the selections from (1,1,1). A subset of these estimates can optionally be selected and added to the existing selections. For example, after examining categories (1,1,0), (1,0,1), (1,0,0) and (0,1,1) using various orderings, it is apparent that some extra selections can be made from category (1,1,0).

Figure 1: SD_State_est_score size versus rank

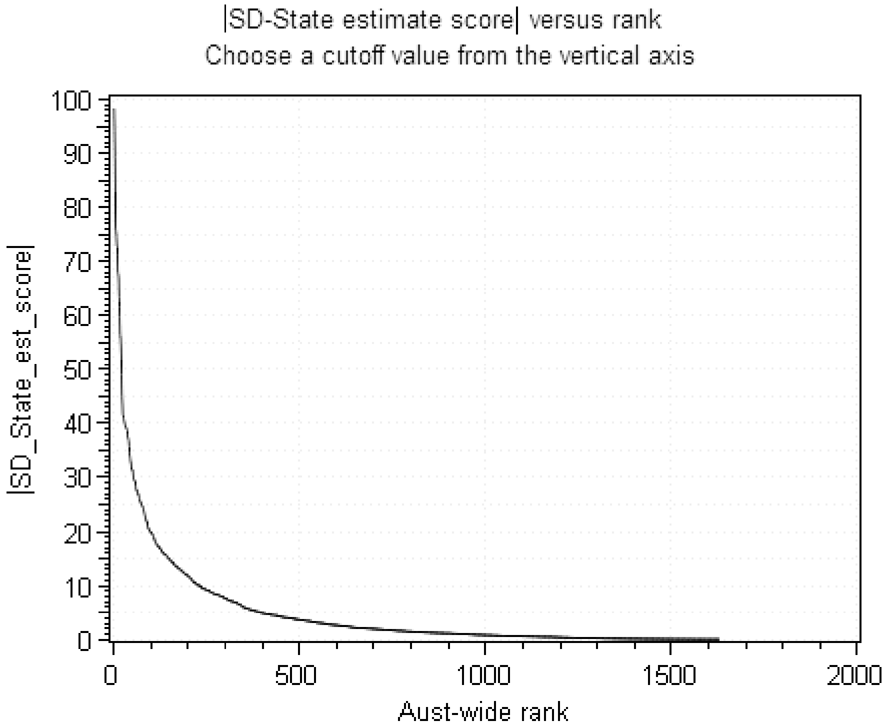

Figure 2: Hierarchical macro edit results for SD estimates

Hierarchical macro editing categories	*Number of SD estimates*	*Number of anomalous estimates*
(0,0,0)	*367*	*.*
(0,0,1)	*407*	*.*
(0,1,0)	*42*	*.*
(0,1,1)	*135*	*.*
(1,0,0)	*61*	*.*
(1,0,1)	*66*	*.*
(1,1,0)	*80*	*.*
(1,1,1)	*493*	*493*
Total	*1,651*	*493*

Cut-offs:

|SD_Aust estimate score| > 0.25
|SD_State estimate score| > 1.75
|SD estimate score| > 15.0

V. Summary and conclusions

A. Discussion

48. Macro significance editing appears to provide a good framework for the general use of scores in macro editing. It has the advantage that it dovetails with the existing micro significance editing framework already in operation in ABS, forming a general significance editing framework. The functionality of the Significance Editing Engine tool can be extended to cover the objective detection component of macro editing. The macro significance scores and sensitivity measures developed within the framework are based on a logical definition of significance which allows for new scores to be developed as the situation demands. The scores make use of macro editor expectations for the data when available. This leads to new scores such as ratio scores which improve on existing ratio scores since they incorporate more information (such as the use of historical estimates for calculating current ratio scores). The framework allows for estimates such as standard errors for sample surveys and coefficients of variation for censuses to be included in scores and for scores to be combined. This allows scores such as macro scores and combined estimate and ratio scores to be developed.

49. Hierarchical scores and macro edits provide very useful tools for addressing swamping and

masking problems and appear to provide good alternatives to Hidiroglou-Berthelot macro edits for both historical and current ratios. Although they make use of macro editor expectations when available, they can be used when expected estimates are not available (and, therefore, remain a viable alternative to Hidiroglou-Berthelot macro edits). They are easy to understand and encourage editors to interact with the data (particularly for movements in estimates). They use explicit, easy to understand, edit boundaries (cut-offs) which allow for flexibility in dealing with conflicting macro editing priorities. They allow macro editor flexibility in choosing anomalous estimates.

50. The significance framework encourages efficient use of editor resources by allowing editing managers to make informed decisions about what to edit and how much to edit. It uses simple interactive cut-offs and supports graphical displays such as graphs of score verses rank and cost/benefit curves which help to visualise the macro editing workload. The use of ranks in the framework is an important element in this regard.

B. References

Australian Bureau of Statistics (2007) *The Editing Guide: The Key to Validating Data from Businesses*, Beta version 4, ABS, Canberra.

Australian Bureau of Statistics (2011, to appear) *Significance Editing Manual*, ABS Corporate Manuals, ABS, Canberra.

Farwell, K. (2004) “The General Application of Significance Editing to Economic Collections”, *Methodology Advisory Committee Papers*, cat. no. 1352.0.55.066, Australian Bureau of Statistics, Canberra.

Farwell, K. (2005) “Significance Editing for a Variety of Survey Situations”, Paper presented at the 55th Session of the International Statistical Institute, Sydney, 5–12 April.

Farwell, K. (2009a) “Hierarchical Macro-editing”, Technical Report, Australian Bureau of Statistics, Canberra.

Farwell, K. (2009b) “Macro Significance Editing Technical Specifications”, Technical Report, Australian Bureau of Statistics, Canberra.

Farwell, K. (2010) “The Use of Scores to Detect and Prioritise Anomalous Estimates”, *Methodology Advisory Committee Papers*, cat. no. 1352.0.55.104, Australian Bureau of Statistics, Canberra.

Farwell, K.; Poole, R. and Carlton, S. (2002) “A Technical Framework for Input Significance Editing”, Conference paper for DataClean2002, Jyvaskyla, Finland.

Gather, U. and Becker, C. (1997) “Outlier Identification and Robust Methods”, in G.S.Maddala and C.R. Rao (eds.), *Handbook of Statistics, Volume 15: Robust Inference*, pp. 123–143, Elsevier, Amsterdam.

Hidiroglou, M.A. and Berthelot, J.M. (1986) “Statistical Editing and Imputation for Periodic Business Surveys”, *Survey Methodology*, 12(1), pp. 73–83.

Maimon, O. and Rokach, L. (eds.) (2005) *The Data Mining and Knowledge Discovery Handbook*, Springer.

Samprit, C.; Hadi, A.S. and Price, B. (1999) *Regression Analysis by Example*, Edition 3, John Wiley and Sons Inc..

Sigman, R.S. (2005) “Statistical Methods Used to Detect Cell-Level and Respondent-Level Outliers in the 2002 Economic Census of the Services Sector”, *Proceedings of the Survey Research Methods Section*, American Statistical Association.