



Methodology Advisory Committee

**26 November 2004
(revised 4 April 2005)**

**The General Application of Significance Editing
to Economic Collections**

Keith Farwell
Statistical Services Branch
Australian Bureau of Statistics
GPO Box 66A
Hobart 7001

email: keith.farwell@abs.gov.au
tel: (03) 6222 5889
fax: (03) 6222 5879

Executive Summary

Selective editing is not new. ABS calls its version, 'significance editing'. It has been used in various guises for the last 10 years in an ad-hoc manner (except for some small groups of economic and agricultural collections). This paper looks to extend the application of significance editing to much more difficult situations. This paper seeks feedback on the approaches taken and instructions on worthwhile future directions.

The paper explains how significance editing has been applied to situations where:

- (i) there are a large number of key variables
- (ii) no data for before editing and after editing has been saved
- (iii) there are one-off surveys and surveys without historical data
- (iv) surveys which are unable to reasonably predict future estimates
- (v) surveys where responses are not able to be predicted suitably
- (vi) surveys have editing strategies that need more than identification of providers.

The paper also looks into best use of significance editing and best alignment of the various editing streams.

Discussion Points for MAC

Main points to centre on are:

- (i) views on the solutions provided
- (ii) views on best utilisation of significance editing
- (iii) ideas for future advancement

Detailed issues involve:

- (i) use of Euclidean and root mean square scores with item weights
- (ii) interactive cutoffs and cost/benefit curves
- (iii) standardised scores
- (iv) dealing with anomalous scores within an interactive approach
- (v) use of GINI indexes
- (vi) use of means and medians to support significance editing
- (vii) significance editing without expected values

The General Application of Significance Editing to Economic Collections

1. Introduction

1 Put simply, input editing can be described as the examination of data for the purposes of error correction (ABS Editing Manual, 1993). Although it is preferable to stop errors before the data has been recorded (Linacre and Trewin, 1993), reporting errors do occur. The aim in input editing is to identify and resolve errors in provider records before any aggregation of data has occurred. The traditional approach had been to run data through a set of edits which identify data considered to be either erroneous or questionable. Editors would then work through the 'edit failures' in an attempt to correct all identified errors. This was a costly and time-consuming process. Many studies have found that there is a tendency to do too much input editing. The material is summarised in Grandquist and Kovar (1997). In fact, for many economic collections, a large proportion of response errors tend to have little collective impact on the final key outputs when corrected (Anderson, 1989; Greenberg and Petkunas, 1986). With limited resources, traditional micro editing techniques are no longer affordable. Various modern editing strategies have been developed to overcome these problems. One approach is to divide the data into that thought to possibly contain influential errors and that which is not. Only the influential group is subjected to traditional editing. The rest may be left as is or subjected to some form of automatic editing. This approach is commonly called 'selective editing'. Latouche and Berthelot (1990, 1992) first explicitly outlined the basic selective editing philosophy. It relies on the premise that some response errors are more important than others and that not all errors need to be corrected. They introduced the idea of using a score function to categorise the data into critical and non-critical streams where only the critical stream is manually edited. Although they suggested several score functions, no explicit framework was provided for developing the score functions. A variety of score functions and selective editing approaches are currently in use around the world. The Australian Bureau of Statistics (ABS) has developed a form of selective editing which is called 'significance editing' (ABS Editing Manual, 1993; Lawrence and McDavitt, 1994; Lawrence and McKenzie, 2000).

2 This paper will provide some background to the development of significance editing in the ABS. It will explore the practical application of significance editing to economic collections. Various practical problems will be outlined and the methodological solutions used will be discussed. Finally, some views on an overall editing strategy are provided.

2. Brief background to the development of significance editing in the ABS

3 Colwell (1990) conducted a review of business editing in ABS and advised that a better approach to cutting back on editing costs may be to incorporate some elements

of output editing into input editing and that input editing could be concentrated on those units that are most important to the survey estimates. After reviewing Colwell's report, Farwell (1991), with reference to Latouche and Berthelot (1990), outlined a method to create scores which target erroneous unit responses which induce, when corrected, important changes to target estimates. A basic significance editing score is a prediction of the change in an estimate due to correcting reporting errors (that is, it is an estimate of the reduction in reporting bias due to editing). If such a score is not possible to approximate, a score which is correlated to the expected reduction in reporting bias should be used. The rest of the paper defined various possible score functions.

4 A series of editing trials were commenced for ABS business surveys in 2002 which were designed to test the extent to which significance editing could be applied to various economic collections. Several surveys are currently using a test version of significance editing while a processing system which includes significance editing is being developed.

3. Significance Editing Basics

5 The most fundamental philosophy in significance-based editing is that if we can predict the impact of editing actions on the results that we are trying to achieve, we will be in the best position regarding what to edit and how much to edit (Farwell and Raine, 2000). Basic significance editing scores need to predict both the likely error in a data value and the impact that correcting it will have on important estimates. The method needs to be consistent with the estimation methodology. The scores are used to create ranks where the highest score has rank 1, the second highest, rank 2, and so on.

6 To set it up, key survey outputs are identified leading to the selection of a set of key data items (referred to simply as 'items' in this paper). An item score is calculated for each key item response and the associated item rankings can be used to generate a prioritised list for each item. Provider scores and ranks are created using the item scores.

7 'Editing benefit' is defined as the absolute value of the relative change induced in the target estimate as a result of editing and, technically, a score is defined as the expected benefit (Farwell, Poole, and Carlton, 2002). The actual change in the item response will not be known until after editing - nor will we know the final value of the target estimate or the values of the final estimation weights since they depend on the make-up of the final set of responses.

3.1 Item Score

8 For an estimate $\hat{Y} = \sum w_i y_{i,uneditd}$ (which can be expressed or approximated as a linear sum of weighted responses) where w_i is the estimation weight and $y_{i,uneditd}$ is the response value for provider i prior to editing, the impact on \hat{Y} due to editing $y_{i,uneditd}$ is $w_i(y_{i,editd} - y_{i,uneditd})$ where $y_{i,editd}$ is the value obtained after editing.

9 For input editing, we need an item score to approximate the size of this impact so we must predict w_i and $y_{i,edited}$. Also, since we may need to combine item scores, we express the impact as a percentage relative to the size of the estimate \hat{Y} (or its standard error) which must also be predicted. This leads to the following item score if we assume the $y_{i,unedited}$ is in error (Farwell, 1991):

$$(1) \quad s_{y,i} = 100 \left| \frac{\hat{w}_i(\hat{y}_i - y_{i,reported})}{\hat{Y}^*} \right|$$

where \hat{w}_i , \hat{y}_i , and \hat{Y}^* are approximations to the final estimation weight, the true response value, and the expected estimate when all responses are available.

10 Significance scores can be derived for estimates that are not just a sum of linear weighted responses. The main requirement is that the actual impact of a change in a data value on the target statistic can at least be approximated and, for editing purposes, that the parameters can be predicted (even if the predictions are relatively rough).

11 In an ideal input editing world, we would like the scores to be estimated independently of the response rate and of other responses. Design weights, possibly adjusted for expected non-response, may be used as approximate estimation weights. For continuing surveys, historical values are generally used for expected values which may be adjusted for expected 'growth'. Previous estimates are usually used as expected estimates (which may also be adjusted for expected growth). As more responses are obtained, expected values may be adjusted to more accurately reflect current behavior.

3.2 Provider Score

12 When there is more than one key item, a provider will have several item scores. If the item scores are used to derive rankings there will be different rankings for different items for the same provider. A single score or ranking for each provider is usually preferred in order to make decisions about which providers to edit. The scores for a group of items can be combined using a metric to produce an overall score. When all the item groups are ultimately combined, the combined score is called a 'provider' score.

13 A provider ranking can be based on the provider scores or the item rankings. Editing can be performed using the ranked item listings or the ranked provider listings or a combination of both. Cutoffs can be applied to the ranked lists, the item scores, or provider scores where those above the relevant cutoffs are selected for editing.

3.3 The 'AWE' approach

14 As mentioned in the introduction, significance editing was first implemented in the ABS in 1992 for the Australian Survey of Average Weekly Earnings (AWE) - a

quarterly survey which published nine wage to employment ratios for various categories of providers. It is an example of the most basic application of significance editing. Five of these ratios were chosen as key outputs at State level and a score for each State ratio was calculated. It used the following score which was developed along the same lines as (1) but targets an estimate of rate:

$$(2) \quad s_{R,i} \approx 100w_i \left| \frac{\hat{w}_i \hat{Z}^* (\hat{y}_i - y_{i,reported}) - \hat{Y}^* (\hat{z}_i - z_{i,reported})}{\hat{Z}^* [\hat{Z}^* - \hat{w}_i (\hat{z}_i - z_{i,reported})]} \right|$$

where $\hat{Z} = \sum w_i z_i$ and $\hat{R} = \hat{Y}/\hat{Z}$ and * indicates approximated estimates.

15 The maximum score of the 5 item scores functioned as a provider score. Providers which had failed the usual micro edits but had a score higher than a prespecified score cutoff were selected for editing. Those responses that passed the edits and those that failed the edits but were below the cutoff were left uncorrected. Historical data were used to provide approximations for the true response value (that is, expected values) and previous estimates were used to approximate estimates.

4. Practical Issues

16 The AWE approach is a simple application of significance editing. It is an easy method to implement when conditions are appropriate. This section will discuss these conditions. Hedlin (2001) provides a useful analysis of this approach.

4.1 Availability of before-and-after data and prespecified cutoffs

17 The AWE approach involves determining a cutoff value prior to receiving responses (called a prespecified score cutoff). Data values before input editing and after input editing (called 'before-and-after' data in this paper) are needed to set up a prespecified score cutoff. Score cutoffs are chosen with the intention that a manageable number of providers will be selected for editing and that a suitable amount of editing benefit will be achieved. The analyst's ability to do this is affected by the volatility of the data and capability to predict key item values and key estimates.

18 Good quality before-and-after data are needed to generate the cutoffs. The data needs to have been intensively edited for it to be of suitable quality, otherwise it is not possible to observe the characteristics of the reporting error distribution. The main advantage of using prespecified cutoffs is that editing can commence as soon as the first response is received. Some disadvantages are that the number of providers selected for editing can vary and the total amount of expected benefit associated with the selected providers can vary. Cutoffs will need to be reviewed occasionally.

4.2 The number of key items and the use of a maximum score

19 The maximum score approach works best when there are a small number of key variables. As the number of key variables increase, the likelihood that a provider record will have at least one item score above the cutoff increases and too many fail. Also, each of the individual item error distributions needs to be analysed since the final cutoff needs to be a compromise between the different optimal item cutoffs. Alternatives to the maximum score are needed when the number of key items increase beyond 5 or 6.

4.3 A continuing survey with stable estimates and historical data

20 The availability of historical data makes creation of expected values a simple and repeatable process. The use of historical values relies on having continuing surveys with a high overlap of selections. The surveys also need to be reasonably stable - that is, a stable sample design and a stable set of variables of interest which are not too volatile. Historical values generally do not need to be adjusted for 'growth' though there could be situations where this is necessary. For example, agriculture commodity values often need adjustment as environmental conditions change from season to season.

5. Methodological issues

21 There are a number of methodological issues that arise when conditions differ from those suitable for an AWE-style significance editing setup. This section will outline our significance editing approach for situations where there is one or a combination of the following: a large number of key items; no before-and-after data; surveys that are run infrequently or are one-off; no historical data; key estimates which are not able to be predicted; no expected estimates; no expected values; and editing strategies that target items rather than providers.

5.1 Combining a large number of key items

22 We have found that a provider score based on either the Euclidean distance or the root mean square (RMS) of the item scores works well with large numbers of items. The following provider score is based on the Euclidean distance and uses item weights (we will simply refer to it as the Euclidean score):

$$(3) \quad s_i = \sqrt{\sum a_{ij} s_{ij}^2}$$

where a_{ij} is a user-defined item weight and s_{ij} is the item score for item j , provider i . The Euclidean score can be converted to an RMS score by setting the item weights equal to the inverse of the number of contributing item scores.

23 Not only does the Euclidean score perform well with a large number of key items, it appears to perform at least as well as the maximum score for small numbers of items. We have used it successfully in economic surveys with key item counts ranging from 4 to 28.

5.2 Example of the use of item weights

24 Item weights are used in (3) to allow manipulation of the Euclidean score. The user may want to make one item more important than another. As mentioned above, they can be used to convert Euclidean scores into RMS scores. The RMS score offers the advantage that its size can be compared directly to those of the constituent item scores. The RMS score is useful when item scores might need to be grouped prior to creation of a provider score. (For simplicity, we will not distinguish between Euclidean and RMS scores in this paper unless necessary.) Also, item weights can be used when there is not a one-to-one correspondence between key outputs and key items as shown in the following example.

25 The quarterly Australian International Investment Survey (IIS) provides an example where item weights are used to account for an implicit key item grouping structure, conversion to RMS scores, and a complex key item to key output relationship. In order to develop the key item list, 28 key outputs were identified and the survey practitioners felt that they fell into the three major groups below. They felt that each group deserved equal attention (that is, each group was considered as equally important).

Group 1: (AIA, FIA) x (DI, PI, Der, OI) x (Tx, CP)

Group 2: (AIA, FIA) x (Debt, Equity) x (Tx, CP)

Group 3: (Income Debits, Income Credits) x (Debt, Equity).

where AIA = Australian Investment Abroad
FIA = Foreign Investment in Australia
DI = Direct Investment
PI = Portfolio Investment
Der = Financial derivatives assets
OI = Other investment assets
CP = Closing position
Tx = Transactions

26 There are 28 key items which generate the 28 key outputs above but there is not a one-to-one correspondence. Groups 1 and 2 share common key items while Group 3 does not. For example, within the category (or domain) AIA by Tx, the data item 'equity capital' contributes to both the DI estimate and the Equity estimate. A reporting error for equity capital will contribute to the item score for DI and the item score for Equity. On the other hand, a reporting error for income will contribute to the income score only. Items weights are needed to account for the fact that several key items contribute to more than one key output when combining scores.

27 Group scores were created and made equally important by converting them to RMS scores as shown :

$$(4) \quad s_{group1} = \sqrt{\frac{s_{group1,1}^2 + \dots + s_{group1,16}^2}{16}}$$

$$(5) \quad s_{group2} = \sqrt{\frac{s_{group2,1}^2 + \dots + s_{group2,8}^2}{8}}$$

$$(6) \quad s_{group3} = \sqrt{\frac{s_{group3,1}^2 + \dots + s_{group3,4}^2}{4}}$$

where the subscripts indicate the 16 items in group 1, 8 in group 2, and 4 in group 3.

28 When combining the RMS group scores to create the provider score, errors in items contributing to Groups 1 and 2 contribute twice to the group scores while errors for income items contribute once. The impact of the income group score is doubled to address this imbalance. The resultant RMS provider score is:

$$(7) \quad s_j = \sqrt{\frac{s_{group1}^2 + s_{group2}^2 + 2s_{group3}^2}{4}} .$$

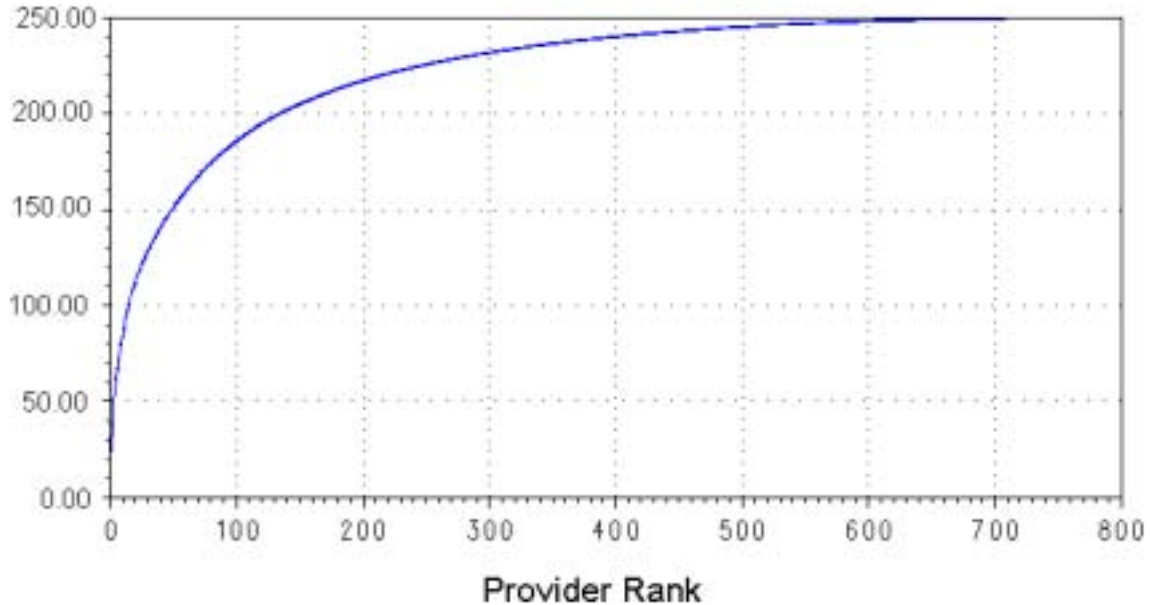
(7) can be defined using (3) with item weights of 1/64, 1/32, and 1/8 for item scores in groups 1, 2, and 3 respectively thus bypassing the need for explicit group scores.

5.3 No before-and-after data

29 As mentioned, if there is no useable before-and-after data, interactive cutoffs may need to be used. Cutoffs can be chosen by reference to cost/benefit graphs (either item graphs or Euclidean provider graphs). These display cumulative expected benefit (or score) against the cost of editing (which is equivalent to plotting by rank when providers have equal editing cost). After editing, if the expected values are replaced by edited values in the scores, achieved benefit graphs can be produced. Graph 1 below is an example of a provider expected cost/benefit curve using the Euclidean score:

Graph 1: Provider cost/benefit graph

Cumulative Expected
Benefit (%)



30 Euclidean and RMS scores provide a more realistic overall measure for the importance of the provider by neatly balancing the various item scores. The score can be roughly interpreted as an average of the item scores which allows a straightforward interpretation of cost/benefit graphs. This makes the Euclidean score very useful for choosing provider cutoffs in an interactive manner which is one of the main advantages of a Euclidean score.

31 Currently, most ABS business surveys do not have before-and-after data and it is getting much harder to obtain such data due to ever-shrinking resources. Survey areas can no longer afford to sufficiently over-edit the data to produce the a before-and-after dataset of suitable quality.

5.4 Cumulative standardised score cutoffs

32 We have found it convenient, with interactive cutoffs, to either augment or replace actual scores (such as those used in the AWE case) with 'standardised' scores. A provider score is standardised as follows:

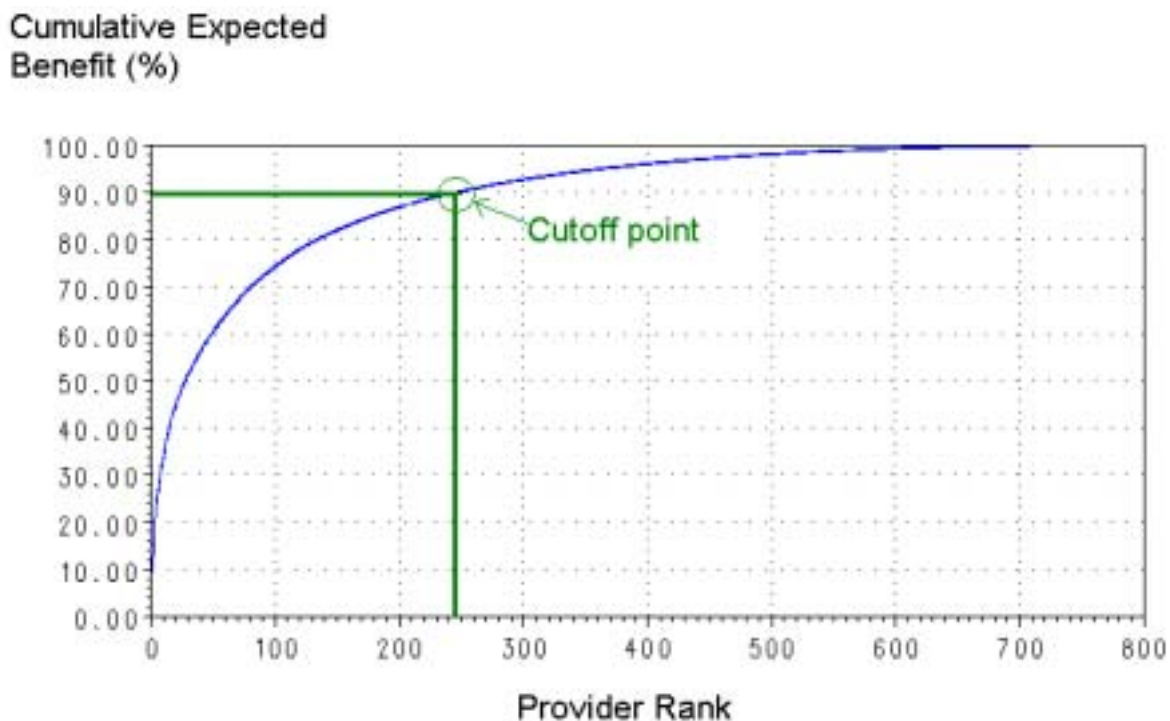
$$(8) \quad s_i^* = 100 * s_i / \sum_i s_i$$

where the sum is over all available providers within the domain of interest. For example, if the key estimates are industry level estimates, then scores are standardised at industry level.

33 A cumulative standardised score is created by aggregating the standardised scores in provider rank order. A cumulative standardised score cutoff involves specifying what percentage of the total possible benefit, or score sum, is desired. This approach works best when there are a suitable number of providers available for editing to ensure that a cost/benefit tradeoff can be achieved.

34 Graph 2 below is the standardised version of Graph 1 and is set up for using a cumulative standardised score cutoff. For example, choosing a 90% cutoff results in editing the top 243 providers available for editing.

Graph 2: Provider cost/benefit graph using standardised scores



35 The number of providers selected for editing can vary between edit runs when using a standardised score cutoff but the amount of expected benefit associated with the selected providers is controlled. We try to decide on a cutoff which we believe will be a good balance between the number of providers we edit and the amount of benefit we expect to achieve. It also allows the user to use a single cutoff across a group of items or domains. The cutoff can be easily adjusted if too many or too few selections result.

5.5 Anomalous scores and an interactive cutoff approach

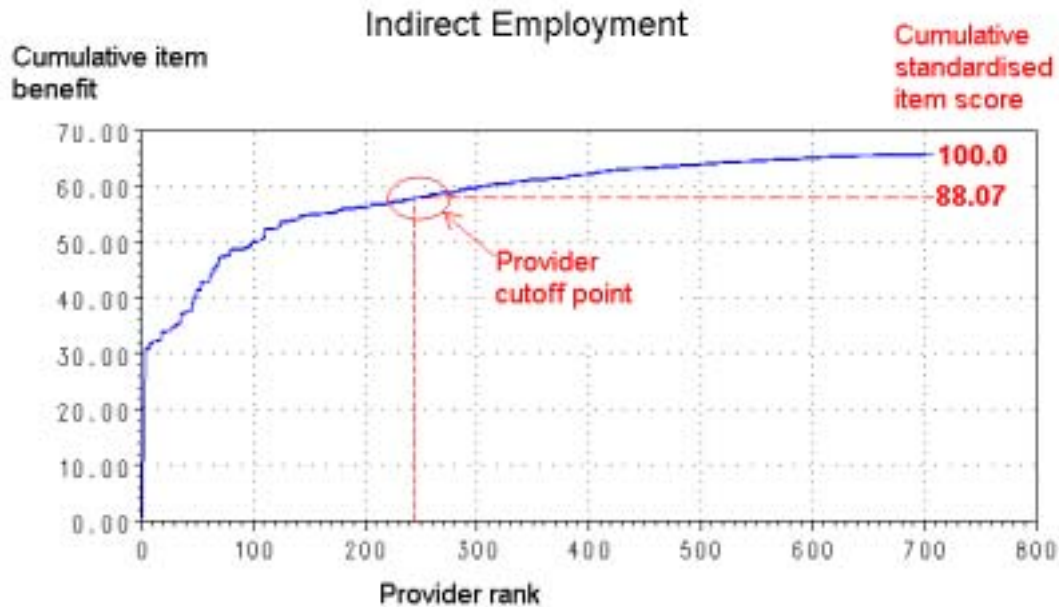
36 When using an interactive approach with standardised scores, there may be a need to adjust for the influence of very large scores since they affect apparent cost/benefit behaviour. Standardised scores are not independent of each other whereas the non-standardised scores are. This paper will refer to very large standardised scores as 'anomalous' scores.

37 One way we have dealt with anomalous scores is to run a 'highest rank' option where we use the highest of the key item ranks as a provider rank and choose a manageable number of the top ranked providers to initially edit. This is a good way to start editing as it quickly identifies any anomalous scores within all key domains and across all key items. The 'highest rank' option is very simple and only requires the specification of how many providers the user is willing to edit. It is very robust against a possible standardised score bias caused by differing numbers of contributors to key estimates. For example, in the Australian Agriculture Survey, some commodities have a huge number of contributors while other rarer commodities have much smaller numbers of contributors. Those commodities with less contributors will tend to generate higher standardised scores on average to those with many more contributors. If the highest rank option was replaced by an option based on using the maximum of the standardised item scores, there would be a tendency to over-select providers that contribute to rare commodities. When there are not problems like those above, score size can be used to select a suitable initial set of anomalous providers. The remaining standardised provider scores are adjusted by removing the impact of the anomalous scores.

5.6 Fine-tuning edit selections with interactive cutoffs

38 When using interactive cutoffs and the number of cost/benefit graphs is not too large, a user might use only item score cutoffs. Alternatively, providers selected using a provider score cutoff can be augmented by choosing extra providers chosen at the item level. This ensures that any major problems with single key items are covered within resource limits at the input editing stage. If using an item cost/benefit curve, there is the choice of which ordering to use - a specific item rank ordering where each cumulative item benefit graph would have its own item ordering or the provider ordering where all item benefit graphs have the same provider ordering. The advantage of provider ordering is that it is easy to make decisions about whether to select extra providers for several key items simultaneously because the total number of selected providers is known immediately. The tradeoff for simplicity is a slightly less efficient gain in expected item benefit. Graph 3 below is an example of an item benefit graph with provider ordering from the Australian Employment Placement and Contract Services Survey (EPCSS). It shows the expected benefit for income from direct employment due to editing the 243 providers chosen using the cutoff shown in Graph 2 (which covers 88.07% of the total item benefit).

Graph 3: EPCSS item cost/benefit graph for income from indirect employment



39 We have examined applying standardised score cutoffs at the item level, a group of items level, and the provider level and tend to prefer the provider level. Cutoffs at the item level require more effort to maintain and may result in editing more providers than feasible for particular items. Generally, a provider level cutoff tends to result in a good balance of expected benefit across the key items while offering the simplest cutoff functionality.

5.7 Large numbers of item cost/benefit graphs

40 Fine-tuning cutoffs can be very difficult when there are too many cost/benefit graphs to manually examine. For example, in the annual Australian Manufacturing Survey, 10 key items were chosen with the key domains being Australia by 151 industries. There were 151 provider cost/benefit graphs and 1510 item cost/benefit graphs. Using a single cumulative standardised score cutoff across all industries can result in some being over-edited and others under-edited. Some cost/benefit curves look very 'curly' and few providers are selected while others are relatively flat and many providers are selected. We might want to alter the cutoffs for the worst cases. GINI coefficients can be calculated for the curves by treating them as special cases of Lorenz curves. The coefficients are essentially an index of curvature and cost/benefit curves can be ordered by them. We used the coefficients to cut down on the number of

graphs to be reviewed by looking only at those with the smallest and largest coefficients and ignoring the rest.

41 We examined the use of provider contributions to the GINI coefficient as a possible tool for use in cutoff methodology. For example, an alternative to an anomaly phase when using the cumulative standardised score cutoff could be developed based on the provider contributions to the GINI coefficient. Although some methods were examined, we were unable to develop a general method that could be methodologically supported. As an example of the methodological issues, we ordered providers by score size removed them one at a time (without replacement) while calculating the GINI coefficient for the cost/benefit curve based on the remaining providers. The new coefficient was associated with the removed provider so that each provider is associated with an adjusted GINI coefficient. We stop selecting providers once the adjusted GINI coefficient is under a chosen GINI cutoff value. However, this can result in a non-monotonic sequence of adjusted coefficients. Further research is needed.

5.8 Rare and one-off surveys

42 Rare and one-off surveys do not tend to have historical data or previous estimates. We have looked at using basic means and medians as expected values for some economic surveys and found that they work reasonably well. The EPCSS had data from 1998/99 which, although the data was considered not suitable for expected values, could be used to develop regression models. We used the current response data to create expected values using hot deck means and medians and using the regression models. We produced a regression-based editing list and an alternative means-based list of ranked providers. The regression-based list was very similar to the mean-based list and either could be used. For example, there were 38 providers in the top 45 common to both listings. The results indicate that significance editing with interactive cutoffs can be used on a survey without historical data and without before-and-after data and that hot deck means and medians can provide useable expected values.

43 Another issue involves difficulty with developing useful expected estimates. One obvious approach is to use the expected values to generate expected estimates. However, our experiences indicate that estimates based on them can be very inaccurate (particularly for those that can be positive or negative). However, they could be used as a starting point. In fact, in the above example, we created 'guesstimates' after looking at estimates from five years earlier, looking at those generated using the current data, and looking at those generated from the expected values. Generally, it is easier to use provider scores based on standardised item scores which are outlined in the following section.

5.9 Unpredictable estimates and standardised item scores

44 Item scores can be standardised as follows:

$$(9) \quad s_{y,i}^* = 100 * s_{y,i} / \sum s_{y,i}$$

where the sum is over all item scores targeting \hat{Y}^* .

45 Standardised item scores do not need expected estimates. Although expected estimates are used to create the original scores, they cancel out in (9). Standardised item scores can be used when the target estimates are either too erratic to predict or not available.

5.10 The IIS example

46 The quarterly IIS, introduced in section 5.2, has extremely erratic estimates that are strongly influenced by very large sparse one-off positive or negative values. Several quarters of before-and-after data were collected for the study and historical data were available.

47 The prespecified score cutoff approach was rejected due to the volatility of IIS estimates. Not only are they extremely difficult to predict but they are also subject to large revisions over several quarters. One major difficulty occurs when estimates began to approach zero - scores become unwieldy. One possible remedy is to manually intervene and alter those expected estimates that seem very poor - but this is far too labour intensive. Another approach could be to create a set of 'standard' expected estimates representing the long term 'average' figures. The preferred methodological approach would be to develop a set of 'bias tolerance' limits in line with specified survey quality requirements. These could be used instead of expected estimates to create the item scores (currently, IIS use the previous quarter's estimates as expected estimates).

48 As an example of the effect of an erratic estimate, the March quarter simulation used the previous foreign investment in Australia equity transaction estimate of \$0.415m as an expected estimate. This was very poor expected value and caused extremely large equity item scores. This item score tended to dominate all other item scores. The Maximum provider score was much more sensitive to this problem than the Euclidean provider score. We looked at using some alternatives for the expected estimate such as \$3.5m and \$3500m. We found that although the system was fairly robust to differences in expected estimate, it was affected when the expected estimates approached zero (as in the \$0.415m case).

49 Ultimately, IIS management wanted a system that was as simple as possible and settled on using standardised item scores with a 98% cumulative standardised provider score cutoff. The system is run at regular intervals and all available unedited providers are put through the system. Providers not selected for editing are put through succeeding runs. With this setup, an anomalous score phase is often not needed. Edit cutoffs are determined so that total expected benefit for the whole quarter is 98% of the total possible expected benefit. This requires keeping a running total of the expected benefit covered by the previous edit selections and adjusting the current score cutoff to maintain the required 98%.

50 We looked at applying this cutoff approach at the item level, a group of items level, and the provider level and chose the provider level. Item cutoffs require more effort to maintain and can cause too many providers to be selected. This was usually due to there being at least one problematic item each quarter. Problem items generate large scores due to dubious expected values that can dominate the provider scores. Using a provider level cutoff balanced the effort across items so that most items attained the required quality after editing while any items that 'slip through' can be tracked down in output editing (even though the desired overall expected editing benefit is achieved in this process). A provider cutoff, therefore, gave greater scope to manage overall editing workload while providing the simplest cutoff functionality. This approach at provider level resulted in requiring about 55% of forms to be edited to achieve about 98% of the benefit that would have been achieved by editing 100% of the forms. The 98% level was also achieved across most key items.

5.10 Significance editing without expected values

51 If it is not possible to obtain expected values, the usual scores cannot be created. In line with a significance-based methodology, we can try to base a score on something that is correlated with the size of reporting errors (Farwell, 1991). For economic variables, providers with important reporting errors are often large contributors to key estimates, their movements, and/or their standard errors. This positive correlation can be used to create a score when expected values are not available (Farwell and Raine, 2001).

52 This method differs from those using expected values by using three 'initial' scores to form an item score. The three initial scores are called the level, movement and standard error scores and they are created by standardising the absolute values of the approximate provider contributions to the level, movement, and standard error of the target estimate.

53 For an unbiased estimator of the form $\hat{Y} = \sum_j w_j y_j$, where w_j is the estimation weight and y_j is the response value for provider i (and the sum is over all responding units), the provider contribution to \hat{Y} is:

$$(10) \quad c_{l,i,y} = w_j y_j$$

and the level score for provider i is:

$$(11) \quad s_{l,i,y}^* = 100 * |c_{l,i,y}| / \sum |c_{l,i,y}|$$

where the sum is over all providers contributing to \hat{Y} .

54 The movement and standard error scores, $s_{m,i,y}^*$ and $s_{se,i,y}^*$, are created in a similar manner. If the movement is estimated as the difference between the two relevant point estimates, the following contribution to movement component is used:

$$(12) \quad c_{m,i,y} = w_{t,i}(y_{t,i} - y_{t-1,i}).$$

55 Noting that $w_{t,i}y_{t,i} - w_{t-1,i}y_{t-1,i} = w_{t,i}(y_{t,i} - y_{t-1,i}) + y_{t-1,i}(w_{t,i} - w_{t-1,i})$, the second term on the right hand side is not used in (12) since it can be considered to represent the component of the contribution due to the change in the estimation weights. We are not interested in this for input editing purposes. However, we are interested in the first term on the left hand side which can be considered to represent the component of the contribution due to the change in reported values.

56 This score can be directly calculated for providers that have responded for both the current and previous time periods but alternatives are needed for new responding providers. One option is to impute a value for $y_{t-1,i}$ and an alternative is to set $s_{m,i,y}^* = 0$ and account for it when the initial scores are combined through the use of the item weights (this will be outlined below).

57 The standard error score is approximated using the square root of the provider contribution to the sample variance of \hat{Y}^* . Some providers may not contribute to the sample variance so we set $s_{se,i,y}^* = 0$, and account for it in the item weights.

58 The initial scores are combined to form an item score as follows:

$$(13) \quad s_{i,j} = \sqrt{w_{l,i} s_{l,i,j}^{*2} + w_{m,i} s_{m,i,j}^{*2} + w_{se,i} s_{se,i,j}^{*2}}$$

where $s_{i,j}$ is the item score for item j , provider i and $w_{l,i}$, $w_{m,i}$ and $w_{se,i}$ are initial item score weights.

59 Assuming that we did not choose to impute for $y_{t-1,i}$ for new units in the movement score, and that we want to use an RMS score, the item weights shown in Table 1 below are used to account for the problems mentioned above in the movement and standard error scores.

Table 1: Item weights for significance editing without expected values

	$w_{l,i}$	$w_{m,i}$	$w_{se,i}$
Continuing providers with non-zero variance contribution	1/3	1/3	1/3
Continuing providers with zero variance contribution	1/3	2/3	0
New providers with non-zero variance contribution	1/3	0	2/3
New providers with non-zero variance contribution	1	0	0

60 Note that, in Table 1, ‘continuing’ refers to providers that are considered to have responded for the item of interest in both time periods and ‘new’ refers to those considered to have responded for the current period only.

61 Having obtained an item score, the rest is similar to the significance editing approaches already discussed.

62 Significance editing without expected values has some advantages over significance editing with expected values. It may be the easiest option for surveys where forming expected values is not possible, for one-off and irregular surveys, or for those survey areas wanting to place a much greater emphasis on output editing.

63 We tested significance editing without expected values for EPCSS. Firstly, we tested the situation where no historical data was available. To do so, we set the movement score and associated item weights to zero. We doubled the item weight of the standard error score in Table 1 for continuing providers with non-zero variance contribution and we tripled the item weight of the level score in Table 1 for new providers with non-zero variance contribution. Secondly, we wanted to try using all available data and, as an alternative, included a movement score by using the 1998/99 data for continuing units and a set of imputes based on it for new units. Item weights were adjusted accordingly. We created 2 provider listings, List 1 and List 2 below, using the same data as was used for the original EPCSS significance editing study:

- (1) List1: Euclidean provider score without movement contribution included;
- (2) List2: as for List 1 except we used the 98/99 data to match providers and created means for non-matched providers to generate pseudo movement scores for new units.

64 We did not have time to do a full analysis which would have included an analysis of relative biases and benefits. However, we did match the providers from each of the above provider lists to the original significance editing lists described earlier in this paper. Table 2 below displays some results.

Table 2: Results from matched edit lists

Original significance editing list	Top 50	Top 100	Top 150	Top 200	Top 250
List 1 common selections	33	65	104	147	180
List 2 common selections	35	67	105	143	179

65 The high match rates between the providers on the significance editing without expected values lists and those on the original significance editing list indicates that we should expect significance editing without expected values to perform reasonably well as a significance-based editing strategy when other alternatives are not available. Some surveys have a far more output-focussed approach to editing and, although they may be able to implement significance editing with expected values, they may decide that the setup is too resource intensive. The now defunct Australian Agricultural Finance Survey used significance editing without expected values as the main editing component very successfully. Initially, it was used as an input editing technique. Then, as publication time got nearer, the functionality was used more as an output editing tool. For example, the level score provided a ranked list for detecting outliers (where impacts on key items could be combined with Euclidean scores). A movement score based on the total contribution was used to tidy up and understand the movements in selected estimates and the standard error score was used to examine the sources of high standard errors.

5.11 Non-form based editing strategies: The Agriculture Survey example

66 ABS run an annual Agriculture Survey (AS) of around 30,000 farms and a quinquennial census. Information on over 600 data items is produced. Estimates for these items are needed at 66 statistical division (SD) levels in survey years and at 1353 statistical local area (SLA) levels in census years. A high proportion of the data items are agricultural commodities and each provider can report for a differing number of commodities. This makes editing strategy needed for the AS different from those needed for the other economic collections mentioned in this paper.

67 'Commodity' collections such as the AS are typically very large collections where providers report on varying numbers of commodities (or items). Although the reported

data on a form tend to form small groups of related items, the item groups themselves are generally not related to each other. The small item groups are important for data verification. For example, the significance editing strategy might identify production of wheat as questionable. An editor would need to also look at the area of wheat reported to appropriately resolve the issue. In this case, area and production form a data item group for editing purposes.

68 For most economic collections, solving an editing problem usually involves looking at the rest of the form since much economic data has a balance sheet structure to it and many items are related. Therefore, form-based editing strategies tend to be used for these collections and editing strategies are based on provider scores and listings.

69 This is not the case for some commodity collections where an item-based (or non-form based) editing strategy is needed. There are many item groups which are unrelated to other item groups. It is not necessary to examine other data on the form to resolve failed item groups. To minimise the editing effort, we not only need to identify the forms needing attention, but also the item groups within the form that need attention. Editors often do not have the time to look over the whole form. Accordingly, provider scores are less useful for the basic editing needs of commodity collections - though they are of use for managing provider re-contact.

70 The approach used here is that a form is selected for editing when at least one key item fails the significance edits. When a key item fails (which may be due to an Australian, State, or SD cutoff as outlined below) the item group is selected for editing. For provider ranking purposes, we used the maximum of the failed item scores as the provider score and order providers by score size.

71 To set up the editing strategy for the AS, we had to identify the key data item groups and key domains. Agriculture believe that information must be focussed on at least the SD level, even though State and Australian aggregates are very important. Key items will differ depending on the SD as different SDs produce different groups of commodities depending on location and environmental conditions. We based the key item list on the survey design variables which are identified each year by examining the value of major commodities reported in the previous cycle at Australia, State, and SD level. There are about 120 key items across all levels. For the 2002/03 cycle there were 49 key items at the Australia level forming 29 item groups, 58 key items for Victoria forming 22 item groups and so on down to a certain number of key items for each of the 66 SDs.

72 We used scores based on historical data and applied cumulative standardised item score cutoffs at Australia, State, and SD levels. To demonstrate the technique, we used a 90% cutoff for all Australian key items, an 80% cutoff for all State key items, and a 60% cutoff for all SD key items. Noting that SDs do not cross State boundaries, we show below the results for Victoria taking Australian and SD requirements into account.

73 The results in Table 3 below show that for the 29 key item groups (formed from 49 key items) at Australia level we would need to edit 25,535 item groups from 13,514 providers across all States to satisfy an Australian cumulative standardised score cutoff of 90% while ignoring the State and SD requirements. There was a average of 3.82 item groups per form and an average of 1.89 groups requiring examination for each form selected for editing. Results are shown for editing for State constraints while ignoring the Australian and SD constraints. Note that there is some overlap in key items at the different levels. Some items which are key at the Australian level may also be key at State level and/or SD level. Finally, Table 3 shows results simultaneously addressing the Australian, State, and SD constraints. For example, there are an average of 4 key item groups per form. 15,311 forms had at least one key item group selected for editing with an average of 2 item groups per failed provider. It is interesting to note that the editing load increased from 25,535 to 30,549 item group failures by adding the State and SD constraints to the Australian constraints (and forms with edit failures rose from 13,514 to 15,311).

Table 3: Edit failure results across all States and SDs

Key domain	Australia	State	SD	Overall
Cutoff value (%)	90	80	60	(90,80,60)
Number of key item groups	29	45	64	65
Number of forms returned	25302	25302	25302	25302
Number of key item group responses	96716	92042	76721	101298
Average number of key item group responses per form	3.82	3.64	3.03	4
Number of forms with key item group edit failures	13514	11457	7902	15311
Number of key item group failures	25535	18642	11183	30549
Average number of key item group failures per failed form	1.89	1.63	1.42	2

74 Table 4 below shows the equivalent results to Table 3 for Victoria. For example, there are 22 key item groups for Victoria using an 80% cutoff ignoring the Australian and SD requirements; the equivalent results for the 34 key item groups across the SDs in Victoria using a 60% cutoff ignoring the Australian and Victorian requirements; and the results for all the 40 key item groups after simultaneously satisfying cutoff requirements at Australian, Victorian and Victorian SD levels.

Table 4: Victorian edit failure results (90% Australian, 80% Vic, and 60% for Vic SD cutoffs)

Key domain	Australia	Vic State	Vic SDs	Vic overall
Cutoff value (%)	90	80	60	(90,80,60)
Number of key item groups	28	22	34	39
Number of forms returned	5168	5168	5168	5168
Number of key item group responses	20817	20231	17451	21480
Average number of key item group responses per form	4.03	3.91	3.38	4.16
Number of forms with key item group edit failures	2744	2018	1685	3008
Number of key item group failures	4439	3018	2309	5550
Average number of key item group failures per failed form	1.62	1.5	1.37	1.85

75 There are various possibilities for deciding on the best set of cutoffs for the above approach. For example, to set the Australian cutoffs, we could look at two graphs which plot the number of item groups and number of forms needing editing against cutoff. Then, for a chosen Australian cutoff, we could produce similar graphs for each State using possible State cutoff choices. The process could be repeated for the SD level using a set of selected State cutoffs (where a single cutoff is applied across all SDs within a given State).

76 The above results indicate the large size of the workload generated by the large number of key items, the multilevel key domains, and the selected cutoff levels. To reduce the workload, either the key item set within selected key domain categories

needs to be reduced or the cutoffs need to be altered. This could be done on a case-by-case basis by manually looking through each item cost/benefit curve which may be too labour-intensive. As an alternative, the GINI coefficient for each key item group by key domain could be used to order the cost/benefit curves from the smallest GINI coefficient to the largest. Editing for items within domains with small coefficient values may be reduced if it is decided that the cost of editing is too high compared to the benefit.

77 Even more basically, it is probably necessary to reassess the need to edit each key domain with the same effort. For example, should the same cutoff be used across all SDs? Also, should we use the same cutoff across all items within each level. It may be beneficial to reduce the cutoff for some Australian level items or maybe some State level items.

78 The above example demonstrates the power of the significance editing framework and how the framework can be used to assist in developing and articulating an editing strategy.

6. General Approach for Setting Up Significance Editing

79 This section will outline the basic steps needed to implement a significance editing strategy. This section should be read in conjunction with Attachment 1 which contains a flowchart of the possible significance editing approaches.

Step 0: Scoping phase

80 This involves working through issues such as the overall editing strategy, the nature of the data, quality needs, timing and available resources.

Step 1: Identify/negotiate key outputs and key domains

81 Domains are the level at which the key outputs are considered important. Some typical examples are State, Industry, State by Industry, etc.

Step 2: Identify a set of key items

82 The key items are chosen from those items which contribute to the key outputs. Auxiliary items may also be identified at this stage. These can be scored and ranked but do not contribute to the provider scores and ranks. They are used to assist editors with error resolution.

Step 3: Create item scores

83 Determine if useable expected values and expected estimates can be obtained.

84 If significance editing with expected values is to be used, create expected values for each selected item. Note that the expected values do not need to be of as high a quality as those needed for estimation purposes. The simplest choice is to use values from an existing item imputation system if one is set up or, if available and considered suitable, historical data. Alternatively, auxiliary data or modeled data may need to be explored. Also, expected values based on the current data can often be used. These can be as simple as using means and medians. Note that the timing of the editing, when using current data to generate expected values, needs to be delayed until there are a sufficient number of responses available.

85 If expected values are available but useable expected estimates are not available, choose between using standardised item scores or using significance editing without expected values. If using standardised scores, decide if anomalous scores need special attention.

86 When using significance editing without expected values, there is a need to decide when best to run it. If it is run before useable estimates have been generated, the standard error contribution needs to be approximated and estimation weights are replaced by selection weights. There is an option to allow expected values to be used (if available) to create pseudo movement scores for providers new to the collection. Within this option, the user can decide to exclude the completely enumerated new providers if their expected values are considered too inaccurate to be useful.

Step 4: Create provider score

87 Is a provider score appropriate? Is the whole form to be examined if it contains edit failures? If the key items on the form are related then a provider score is usually appropriate since resolving the errors usually requires a study of other items on the form. If the key items on a form are unrelated or if there can be different numbers of key items reported on different forms, an explicit provider score may not be very useable. Should items on the form be grouped for scoring purposes?

88 If a provider score is needed, decide on the appropriate scoring and ranking method. Provider score options:

- (i) maximum score of item scores (maximum method);
- (ii) Euclidean or RMS score with optional item groups (Euclidean method).

89 Provider ranking options:

- (i) ranks based on maximum score;
- (ii) ranks based on Euclidean or RMS score;
- (iii) ranks based on highest rank of the item ranks

90 If only selected items or item groups are to be examined, it may still be useful to choose a suitable provider ranking method so that provider contact can be appropriately managed.

Step 5: Define cutoffs and workload

91 For significance editing with expected values, decide if predetermined cutoffs can be developed. If not, or if the method is considered unsuitable, use interactive cutoffs. For significance editing without expected values, interactive cutoffs are generally used.

Step 6: Run system

92 If using interactive cutoffs with standardised scores, it is usually best to initially run the highest rank option (as outlined in section 5.5) or choose anomalies based on highest scores (unless there will be many runs where the unedited providers are put back through and re-scored). Later runs usually use either the maximum or Euclidean provider score (with item groups and item weights incorporated if needed). It is possible to have different runs targeting different domains.

Step 7: Occasionally monitor effectiveness of predetermined cutoffs

93 There may be a need to check for cumulative effects from time to time. Select a sample of providers with repeated below-cutoff scores. For example, some providers might be copying the previous values and the score will be zero if historical values used in the scoring. It may sometimes be desirable to select a sample of providers below the cutoff and edit them to check on the effectiveness of the prespecified cutoffs used.

Step 8: Review process

94 Achieved benefit graphs (when using Euclidean scores) and graphs showing estimate change ordered by provider ranks can be generated to assess effectiveness of the process. Before and after data must be saved.

7. Towards an overall editing strategy

95 Significance editing is only one part of an editing system. Its most basic use is as a prioritisation and ranking device (to facilitate selective editing) although it can also supersede other edits within the system. Many of the conceptual and practical difficulties faced when implementing significance editing are tied up with the total editing strategy. It is a difficult and complex job to compile a suitable editing strategy that deals with the many competing quality and timeliness demands.

96 A selective editing phase can be included in an editing system in several ways. For example, in the AWE approach, significance editing was superimposed on an existing micro editing system. The micro editing system was used to detect the questionable providers while the significance editing system was used to divide the edit failures into critical and non-critical streams. The approach was dependent on the quality of the existing editing system - some micro editing systems might be out-of-date, some might fail almost every record, or the edits might fail too many reported data items within a provider record. There will be a need to maintain both the existing micro editing system as well as the significance editing system. An alternative approach is to use the significance editing system to detect the questionable providers (or items) and use micro edits to assist with the error resolution phase.

97 It is possible to develop a very sophisticated significance editing system that merges selective editing with the more traditional micro editing. There is no room to explore this idea here except to say that it involves modifying micro edits so that they become significance edits. For example, the midpoint of a range edit can be used as an expected value and a significance score can be produced which could replace the range edit.

98 The significance editing system can be used as a stand-alone input editing system. The item scores and ranks can be used to assist with error resolution. Example 1 below shows some output displaying a SIS provider ranked 2 and associated detailed key item scores and ranks (all data has been altered for confidentiality reasons).

Example 1: A detailed provider listing

Providers to be Edited to Achieve 90% of Total Editing Benefit (maximum rank = 710)

Rank=2 UNITID=MU00451896 Stratum=30095 Selection Weight=3.33 New Unit?=Y					
Item Description	Reported Value	Expected Value	Item Benefit (%)	Rank	
DEMPINDP - Indirect Employment	31000.0	1988.0	19.80	1	
DLABCOST - Labour Costs	117798000.0	213674670.0	3.54	2	
DINCPERM - Income from Permanent Placements	2449000.0	6585096.0	1.99	8	
DINCTEMP - Income from Temporary Placements	144523000.0	125248622.0	1.04	2	
DDIREMPT - Direct Employment	224.0	121.0	0.76	23	
NUMTEMJP - Temporary Job Placements	30000.0	16527.0	0.31	61	
NUMPERJP - Permanent Job Placements	130.0	683.0	0.30	31	
DTOTEXP - Total Expenditure	315624000.0	145721308.0	0.26	7	

99 We have run significance editing on several varied survey data files and found that significance editing was able to detect questionable providers and items very well. Often, problems in the key items led to non-key items being examined when form-based editing was used. The pertinent issue concerns what is the best way to facilitate error resolution?

100 Based on our experiences to date, we finish this section by listing some views on input editing systems.

6.1 Initial edits

101 These result in what are often called 'fatal' or 'hard' edit failures. For example, simple checks for logical errors which may include code legality checks, consistency checks and logical cohesion checks; 'thousands' edits (and other problems with reporting in wrong measurement units); selected balance edits (such as checks that a reported total equals its components); code combination checks; checks for missing values; checks for rows of 9's (numbers inserted at data entry for various reasons to indicate further attention needed); or even series of 1's in reported data (typical in systematic optical character recognition errors). It seems best that many of these easily detectable errors get corrected as soon as possible so they do not confound things further down the processing path.

102 For significance editing, it is at least necessary to ensure that the data needed to target the scores is correct. For example, if we target State estimates we need to ensure that the State code is valid and consistent with the data. Otherwise, the scoring system will be compromised. If possible, initial edits and their resolution should be automated.

6.2 Micro edits

103 These are edits such as range edits, ratio edits, percent contribution edits, and the like. It is possible that many of these can be eliminated through smart design of edit rules and a well thought-through significance editing strategy. However, for categorical or social data, these kind of edits are still required. Also, micro edits can assist editors to resolve errors even where significance editing is the main editing approach.

6.3 Significance edits

104 Significance edits can be used as a form of selective editing to decide between what goes into a critical editing stream (which will require special attention) and what goes into a non-critical stream (which might be left unexamined or subjected to auto-correction). They can also be used to minimise the information needed from micro edits by assisting in the error resolution process. In fact, scores for non-key items (which are not used to create provider scores) can augment the significance editing outputs.

105 Significance edits can also be utilised in the initial editing phase to deal with special cases such as balance edits. Scores based on the weighted difference between reported totals and the associated derived totals can be used to place some in a critical stream requiring special attention while the rest can be automatically corrected.

106 Although the significance editing approach is very cost efficient and useful, we might also be interested in assuring a certain level of internal consistency within records (such as when we intend to disseminate unit record files or use them for various scientific analyses). It is feasible that the application of a set of micro edits could be of use in the overall editing strategy.

6.4 Macro editing

107 The input editing strategy needs to take account of the macro editing strategy and vice-versa. Aside from the obvious uses such as detecting outliers, analysing data movements, performing exploratory data analysis, it is needed to at least assure the quality of relationships in the data and to safeguard against influential errors slipping through an 'extremely efficient' significance editing phase. It can be used to assure the quality of non-key estimates which are not explicitly targeted in the significance editing stream. These systems lead naturally to the use of graphical tools.

6.5 Auto-correction

108 A sophisticated editing system should utilise autocorrection of edit failures where possible. Some agencies do not attempt to edit the failures from the non-critical stream while others do. For those that do, it is desirable to minimise effort needed and it appears that the best practice approach is the 'Felligi-Holt' method (Felligi and Holt, 1976). Only a few agencies have managed to suitably implement this approach as it is a difficult procedure to implement but is useful within an auto-correction strategy. This approach would be useful for surveys with large numbers of logical edits or data items that are not conducive to significance editing (such as responses to tick-box questions).

6.6 Ideal editing strategy

109 The ideal editing strategy is an intelligent combination of all these possibilities. One aspect that significance editing is very strong on is its general framework which assists users to develop their strategy and philosophy. It is also strong on feedback measures assessing editing efficiency. Ultimately, a sound editing system needs to have a review capability and the flexibility to be modified.

References

Anderson, K. (1989), "Enhancing Clerical Cost-Effectiveness in the Average Weekly Earnings Statistical Collection", *unpublished report*, Australian Bureau of Statistics.

ABS Editing Manual (1993), Australian Bureau of Statistics.

Colwell, J. (1990), "A Review of Editing", *unpublished report*, Australian Bureau of Statistics.

Farwell, K. (1991, revised 2002), "Fundamentals and Suggestions for Ranking and Scoring Edited Unit Responses", *unpublished report*, Australian Bureau of Statistics.

Farwell, K., Poole, R. and Carlton, S. (2002) "A Technical Framework for Input Significance Editing," *DataClean2002 Conference*, Jyvaskyla, Finland.

Farwell, K. and Raine, M. (2000), "Some Current Approaches to Editing in the ABS", *International Conference on Establishment Surveys II*, Buffalo, USA.

Fellegi, I.P. and Holt, D. (1976), "A Systematic Approach to Automatic Edit and Imputation", *Journal of the American Statistical Association*, **71**, 17-35.

Grandquist, L. and Kovar J.G. (1997), "Editing of Survey Data: How Much is Enough?", In *Survey measurement and Process Quality*, eds. L. Lyberg, P. Biemer, M. Collins, E. De Leeuw, C. Dippo, N. Schwarz, and D. Trewin, New York: John Wiley and Sons, 415-435.

Greenberg, B. and Petkunas, T. (1986), "An Evaluation of Edit and Imputation Procedures Used in the 1982 Economic Censuses in Business Division", Bureau of the Census, US Department of Commerce, SRD Research Report Number: Census/SRD/RR-86/04.

Hedlin, D. (2001), "Score Functions to Reduce Business Survey Editing at the U.K. Office of National Statistics", *Journal of Official Statistics*, **19:2**, 177-199.

Latouche, M. and Berthelot, J.-M. (1990), "Use of a Score Function For Error Correction in Business Surveys at Statistics Canada", presented at the *International Conference on Measurement Errors in Surveys*, November 1990.

Latouche, M. and Berthelot, J. (1992), "Use of a Score Function to Prioritize and Limit Recontacts in Editing Business Surveys", *Journal of Official Statistics*, Vol.8 No.3, 389-400.

Linacre, S.J. and Trewin, D.J. (1993), "Total Survey Design - Application to a Collection of the Construction Industry", *Journal of Official Statistics*, **9**, 611-621.

Lawrence, D. and McDavitt, C. (1994), "Significance Editing in the Australian Survey of Average Weekly Earnings", *Journal of Official Statistics*, **10:4**, pp. 437-447.

Lawrence, D. and McKenzie, R (2000), "The General Application of Significance Editing", *Journal of Official Statistics*, **16**, 243-253.

ATTACHMENT 1: Significance editing framework

110 Our experiences to date indicate that a group score level needs to be introduced to the framework. This is not shown in the current flowchart but would be an intermediate step between the item score and the provider score. That is, group scores are created by combining item scores and the provider score would be created by combining group scores.

111 Paths A and B can be used for significance editing with expected values. Paths C and D can be used for significance editing without expected values. Path D can also be used for rudimentary output editing.

112 It is assumed that most elements of the flowchart are self-explanatory and, for simplicity, some details are not included.

