SELEKT – A Generic Tool for Selective Editing

Anders Norberg[1]

The aim of selective editing is to make the often resource-demanding traditional editing process in business surveys more effective without a substantial loss in the precision of the output statistics. Recently, Statistics Sweden has developed a generic software package for selective editing called SELEKT. The method underpinning the software promotes continuous measurement of the suspicion of error response rather than a dichotomous measure using traditional edits. SELEKT is flexible and can be used in the production of surveys with different designs. Business surveys have diverse output regarding the number of variables, statistical measures, and domains of study definitions. The key objective of selective editing is to rank suspected errors in data according to the anticipated impact on the output. The software therefore has options to set different weights for different parts of the output to meet the needs of the main users of the statistics. Statistics Sweden has implemented SELEKT in eleven surveys to date. The experience gained will be used to provide recommendations on how to perform selective editing. This article will give an insight into SELEKT and its underlying theoretical base.

Key words: Edit rule; anticipated value; suspicion; impact; score function.

1. Introduction

The aim of statistical data editing is to detect and adjust errors in data resulting from collection and processing. It is considered a necessary survey operation because errors in survey data may distort estimates, complicate further processing, and decrease provider and user confidence (Granquist 1984). The emphasis of the editing task is slowly moving from just cleaning up the data, though this remains a necessary operation, to identifying and collecting data on errors, problem areas, and the causes of error to provide a basis for a continuous improvement of the whole survey vehicle (Granquist 1997). A further goal is to provide information for the quality declaration of the output statistics.

Editing appears in various forms in the business survey process, from respondents entering information into electronic questionnaires to the final checking of results prior to publication. Methods and procedures are mainly divided into micro- and macroediting. Microediting means the checking of individual data records, preferably as soon as data is

available. Checking rules are logical conditions or restrictions applied in order to check the validity, internal consistency, and plausibility of individual units' data. A routine flags data that fail the checking rule and enables the analyst, who performs the manual editing, to change data interactively. The analyst must have knowledge of the survey, the population and the kind of errors that are likely to occur. The flagged data can be compared to reference data such as data on the same unit from previous survey rounds, data on similar units, and data from an external register or information on the internet. Finally, the editor may recontact a respondent to check whether a value is declared as correct or to obtain a new value for a variable that was originally incorrect or was suspected to be incorrect. Procedures like this are called manual editing, interactive editing, production editing or simply microediting. Macroediting is performed when all, or most of, the data has been collected. The so-called distribution method checks for outliers in the final data set, as microediting does, but now the comparisons with similar units are more in focus. The aggregation method analyses statistical output, in order to check that no influential errors in microdata remain and that no processing error has been introduced.

Particularly in business surveys, editing is recognised as a time- and resource-consuming survey process, especially when recontacts are necessary. The costs include not only financial and human resources, but also loss in timeliness and excessive respondent burden. Further, there is a danger of distorting true values to fit them to preconceived models. This "overediting" gives users a false sense of security as far as data quality is concerned (Granquist and Kovar 1997).

New theories and methods to reduce the resources spent on editing survey data have been developed over the last thirty years or so. The leading idea is to concentrate resources in microediting on observations that affect the estimates, accepting that final data sets do contain errors with a negligible effect on the statistics produced (Granquist and Kovar 1997). Several methods have been implemented for this purpose. An early method, proven successful by experience, was presented by Hidiroglou and Berthelot (1986). This method considers both the level and the relative change from a previous survey round for a survey variable. Robustness is achieved by using the median and quartiles in the analysis of data.

In the early 1990s, methods based on a score function for selective data editing, henceforth abbreviated SE, emerged. In these methods, survey units that fail at least one edit rule are ranked by the score in order to give priority to those units that have the largest overall anticipated effect on the statistics produced. SE yields a global score for a primary sampling unit, for all the data delivered for that unit, that is, for any cluster elements and for one or many variables. The purpose of SE is merely to reduce the cost of the manual follow-up work without a substantial loss in the precision of the output statistics. Reducing manual follow-up with recontacts also lightens the workload of respondents. Macroediting by the aggregate method and SE have apparent similarities in that both methods focus on the set of estimates to be published. It turns out that early manual editing by SE can reduce the late macroediting.

De Waal et al. (2011) emphasise the detection and correction of systematic errors as a first step in an editing process. A systematic error occurs frequently between responding units when they misunderstand or misread a survey question in the same way. Causes of major systematic errors can be discovered through analysis of edit failures in data; frequent failures of an edit rule are indications of a problem for the respondents.

Data values with small systematic errors are difficult to distinguish from true values as they lie in the interior of a statistical distribution. Causes of these so-called inliers must be found by other methods than editing and the problems must be solved by improved data collection instruments. SE methods, leaving a part of data without manual follow-up, are not appropriate when the survey suffers from severe systematic errors.

The Australian Bureau of Statistics, ABS, has developed a form of selective editing called "significance editing", discussed in Latouche and Berthelot (1992), Lawrence and McDavitt (1994), Farwell and Raine (2000), Lawrence and McKenzie (2000), and Farwell (2004). A basic significance editing score is a prediction of the change in an estimate due to correcting reporting errors, as it is an estimate of the reduction in reporting bias. 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 approach has resulted in a noticeable improvement in editing efficiency. ABS has developed the tool Significance Editing Engine, SigEE, presented by Brinkley et al. (2011).

Di Zio and Guarnera (2013) present the generic tool Selemix, developed by the Italian National Institute of Statistics, Istat. They assume a normal model for the true data, also possible in log scale, and an "intermittent" error mechanism such that a proportion of data is contaminated by an additive Gaussian error. Based on these assumptions, a latent class model is used to derive the distribution of "true" data conditional on observed data. This approach allows scores to be interpreted as anticipated impact of errors and allows a selective editing procedure to be defined that identifies units containing the errors that have the largest influence on the estimates of interest.

A series of projects was started at Statistics Sweden, SCB, in 2006 with the main purpose of analysing which methods should be recommended for editing and for construction of a generic IT tool. Case studies focused on how to use SE (Adolfsson and Gidlund 2008). Nine of the most editing-intensive surveys were included in the project. The case studies show that SE will lead to efficiency gains and likely cost reductions in many surveys. The implementation of SE demands intensive testing in every specific survey. The variation between the surveys regarding survey design, data structure, output statistics, and so on, is large. A generic tool for editing must therefore be very flexible to deal with these differences. Efficiency can also be increased by dealing with known or encountered measurement problems in the auditing and picking low-hanging fruit that would improve the surveys.

The next editing project documented the best methods used in the case studies for various situations. Key algorithms for scores and the aggregation thereof were the framework.

The third phase was developing an IT tool. SELEKT is SAS® application for SE which establishes a general solution that can be implemented in many different surveys. The generic approach implies a set of parameters to be set instead of writing code.

This article will give an insight into the theoretical base for SELEKT. In the next section SE is described, both generally and specifically for SELEKT. Section 3 includes methods for computation of anticipated values and intervals of normal variation from background data. Two major views on how to set the limit for the manual editing by SE are presented in Section 4. Finally, experience from implementation and running of SE at Statistics Sweden is summarised in Section 5.

2.   Selective Editing

Suspicion of a data value being in error and the potential impact of a suspected error on output statistics are the two aspects to consider in the search for influential errors. SELEKT produces local scores based on indicators of both suspicion and impact for all variables, all statistical measures, and all domains of study that are considered important. These local scores are aggregated up to a global score for the respondent.

Example 1: Jäder and Norberg (2006) compute a score function for the International Trade in Goods statistics (ITG). Suspicion is based on observed price per quantity for a transaction, relative to normal variation in product groups, countries and direction (imports/exports). Potential impacts on estimates for domains of study are measured as the difference between reported value and anticipated value of trade, relative to the normal size of the domains. In this survey, suspicion and potential impact have a very low correlation and both of these dimensions are important.

Suspicion indicated by traditional edit rules is described in Subsection 2.1 along with alternative and supplementary methods to indicate suspicion. In Subsection 2.2, classes of domains of study in the output statistics are established and impact of errors is defined. In Subsection 2.3, suspicion and impact are combined and local scores are defined. These local scores are then aggregated to the global score in Subsection 2.4.

2.1.   Suspicion

An edit, also known as an edit rule or a checking rule, can indicate that a data value is or might be in error. A fatal edit manifests if a data item is in error. Examples of fatal errors are inconsistent responses, invalid entries, and item nonresponse. Query edits point to suspicious data items. An example could be a value that, compared to historical data, seems suspiciously high. SELEKT makes use of indicators of the degree of suspicion, not only the traditional dichotomous results "accept" and "fail". Three options are available for assigning a level of suspicion with SELEKT: use of traditional edits, use of hit rates for edits, and the SELEKT-type edits. These are presented in Subsections 2.1.1, 2.1.2, and 2.1.3 respectively.
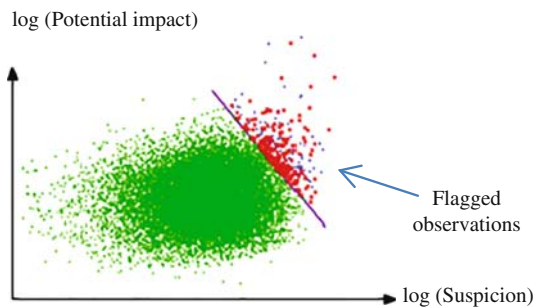

Fig. 1.   *International Trade in Goods statistics. Suspicion is based on price per quantity; potential impact reflects value of trade.*

2.1.1. Traditional Edits

An edit is a logical condition or a restriction on a variable value or a group of variable values which must be met if the data is to be considered acceptable. In the absence of a conceptual model for edits, Norberg (2011) proposes the notion of a *test variable*, which is a function of one or more survey variables, possibly including register variables. The test variable can be the survey variable as is. A common test variable is the ratio of the survey variable to the corresponding variable in a previous survey round. Comparing received data to register information is another example. The value of the test variable is compared with an *acceptance region*, which is the range (or set) of acceptable values for the test variable. The comparison is often made for subgroups of data that are homogeneous with respect to the test variable so that the acceptance regions will be tight and thereby the edit will be efficient. These so-called *edit groups* are defined by using auxiliary variables such as stratification variables or variables used to define domains of study, but need not only be these two types.

Example 2: Say that the survey variables Number of employees and Turnover are measured for enterprises in different industries in a survey. The test variable Turnover/Number of employees is tested with different acceptance regions for three sets of industries. The name of the edit is A01.

Test variable: $z = $ Turnover / Number of employees
Edit groups:
- Industry $=$ M (Manufacture)
- Industry $=$ W (Wholesale trade)
- Industry $=$ R (Retail trade)

Acceptance regions:
- [257€, 550€] for Industry $=$ M
- [505€, 1082€] for Industry $=$ W
- [227€, 632€] for Industry $=$ R.

Edit rule A01 in the software code is:
if Industry $=$ 'M' and not $(257 < z < 550)$
or Industry $=$ 'W' and not $(505 < z < 1082)$
or Industry $=$ 'R' and not $(227 < z < 632)$
then Errcode_A01 $=$ 'Fail'

A failed query edit like this does not necessarily imply that any of the observed, unedited values of the two survey variables are in error. In the example, it may well be the Industry code. The conceptual model, regardless of any lack of universality, is used to construct implicit edits by SELEKT, which is described in Subsection 2.1.3.

At Statistics Sweden, the implementations of SELEKT primarily make use of the existing traditional edits, simply for practical reasons. Suspicion is set to one if the edit is failed and to zero if the data is accepted.

2.1.2. Traditional Edits and the Use of Hit Rates

The dichotomisation of suspicion by 0 / 1 or "Accept"/"Fail" entails a loss of information compared to using varying levels of suspicion. For example, the level of suspicion by a

traditional edit can be set equal to the hit rate of the edit, based on evaluations of previous survey rounds. UNECE (2000) defines:

"Hit rate is the success rate of an edit; the proportion of error flags that the edit generates which generate a change of data in the follow-up process".

Example 3 (cont. from Example 2): Assume that the hit rate for edit rule A01 is known from past survey rounds to be around 70%. Now add the software script: if Errcode_A01 = 'Fail' then Suspicion_A01 = 0.7

When a hit rate is less than 0.8, say, the original edit can be made manifold to utilise information.

Example 4 (cont. from Example 3): When the hit rate for edit A01 is known to be 70%, evaluate previous survey data to find out which wider acceptance regions would have yielded a hit rate of around 90%, say. Excluding the data outside this second acceptance region but keeping data that were flagged by the first acceptance region, the hit rate will be less than 0.7, say 0.6. Let us now make two edits A01a and A01b with the following script.

```
if Industry  =  'M' and not (257 < z < 550)
or Industry  =  'W' and not (505 < z < 1082)
or Industry  =  'R' and not (227 < z < 632)
then Errcode_A01a =  'Fail'
if Industry  =  'M' and not (182 < z < 697)
or Industry  =  'W' and not (360 < z < 1370)
or Industry  =  'R' and not (152 < z < 948)
then Errcode_A01b =  'Fail'
if Errcode_A01a =  'Fail' then Suspicion_A01  =  0.6
if Errcode_A01b =  'Fail' then Suspicion_A01  =  0.9
```

Varying the level of suspicion has a direct proportional effect on the score and thereby makes a difference in terms of the priority given to observations far out from the other observations.

Fatal errors could also be treated in SE, such that the suspicion is set to one. SELEKT has an option to send all fatal errors to follow-up or only those that have a major impact on output statistics. The rest can be imputed in a later stage of the production process.

2.1.3.   SELEKT-Type Edits

SELEKT has a module which constructs implicit edits. Such edits can be used as a complement to or substitute for the traditional edits described above. The user specifies an edit by a test variable. A test variable is an existing variable or an expression of variables in the dataset to be edited, as for traditional edits. Edit groups are automatically defined by SELEKT according to a few specifications by the user. These groups are also used to compute anticipated values, see Section 3. Implicit acceptance regions are computed by SELEKT based on intervals of variation estimated from previous survey rounds.

Tukey's exploratory data analysis, EDA, is an approach to analysing data sets to summarise their main characteristics, often with visual methods. The box plot, based on quartiles of a distribution, has been a basis for the development of the SELEKT-type edits.

In this respect, SELEKT has similarities with the Hidiroglou-Berthelot method. Unlike the Italian Selemix, there is no assumption of any explicit distribution of the errors.

Let k,l identify observed unit (element) l belonging to primary sampled unit (cluster) k. Unit k,l implies data on two levels, say an enterprise delivers data for all employees or all products. One-stage sampling is just a special case with this notation. The notation is also valid for two-phase sampling.

Assume that quartiles and medians are preferred as the basis for the edits. Alternatives are presented at the end of this subsection. Each observed unit k,l belongs to one and only one edit group g. For a set of data from previous survey rounds, let

$\tilde{z}_{i,g}^{L} =$ lower quartile of the *i:th* edited test variable values for the edit group g

$\tilde{z}_{i,g}^{U} =$ upper quartile of the *i:th* edited test variable values for the edit group g

$\tilde{z}_{i,g}^{M} =$ median of the *i:th* edited test variable values for the edit group g.

Let

$z_{i,k,l} =$ unedited value of the *i:th* test variable in the current, unedited data.

The parameter $\kappa > 0$ defines the "gap" of the value range where suspicion shall be zero, that is, the acceptance region. A small κ yields suspicions larger than zero already at a small deviation from the anticipated value $\tilde{z}_{i,g}^{M}$.

The parameter $\tau \geq 0$ defines the regions where suspicion grows from 0 to 1. With $\tau = 0$ the suspicion equals 1 outside the acceptance region defined by κ.

Definition of suspicion $\xi_{i,k,l}$ by *i:th* test variable $z_{i,k,l}$:

If $z_{i,k,l} \leq \tilde{z}_{i,g}^{M} - (\kappa + \tau) \cdot \left(\tilde{z}_{i,g}^{M} - \tilde{z}_{i,g}^{L} \right)$ then $\xi_{i,k,l} = 1$

if $\tilde{z}_{i,g}^{M} - (\kappa + \tau) \cdot \left(\tilde{z}_{i,g}^{M} - \tilde{z}_{i,g}^{L} \right) < z_{i,k,l} < \tilde{z}_{i,g}^{M} - \kappa \cdot \left(\tilde{z}_{i,g}^{M} - \tilde{z}_{i,g}^{L} \right)$ then

$$\xi_{i,k,l} = \frac{\tilde{z}_{i,g}^{M} - \kappa \cdot \left(\tilde{z}_{i,g}^{M} - \tilde{z}_{i,g}^{L} \right) - z_{i,k,l}}{\tau \cdot \left(\tilde{z}_{i,g}^{M} - \tilde{z}_{i,g}^{L} \right)}$$

if $\tilde{z}_{i,g}^{M} - \kappa \cdot \left(\tilde{z}_{i,g}^{M} - \tilde{z}_{i,g}^{L} \right) \leq z_{i,k,l} \leq \tilde{z}_{i,g}^{M} + \kappa \cdot \left(\tilde{z}_{i,g}^{U} - \tilde{z}_{i,g}^{M} \right)$ then $\xi_{i,k,l} = 0$

if $\tilde{z}_{i,g}^{M} + \kappa \cdot \left(\tilde{z}_{i,g}^{U} - \tilde{z}_{i,g}^{M} \right) < z_{i,k,l} < \tilde{z}_{i,g}^{M} + (\kappa + \tau) \cdot \left(\tilde{z}_{i,g}^{U} - \tilde{z}_{i,g}^{M} \right)$ then

$$\xi_{i,k,l} = \frac{z_{i,k,l} - \left(\tilde{z}_{i,g}^{M} + \kappa \cdot \left(\tilde{z}_{i,g}^{U} - \tilde{z}_{i,g}^{M} \right) \right)}{\tau \cdot \left(\tilde{z}_{i,g}^{U} - \tilde{z}_{i,g}^{M} \right)}$$

if $z_{i,k,l} \geq \tilde{z}_{i,g}^{M} + (\kappa + \tau) \cdot \left(\tilde{z}_{i,g}^{U} - \tilde{z}_{i,g}^{M} \right)$ then $\xi_{i,k,l} = 1$

When the lower quartile equals the median, the suspicion $\xi_{i,k,l}$ equals 0 for observed values lower than the median and analogously for the right side of the distribution. This could of course be considered to be a problem, so a manual intervention is recommended. SELEKT gives easy access to the files of quartiles.

The suspicion function is illustrated in Figure 2. Notice that this function considers a skewed population by an asymmetric acceptance region. Generally, a logarithmic transformation for skewed test variables is also recommended.
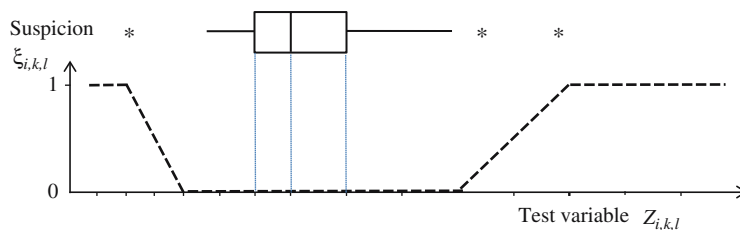
Fig. 2.   Illustration of the SELEKT*-type method to measure suspicion. Assume that the test variable in historical edited data is distributed as shown by the box plot. With the parameters $\kappa = 3$ and $\tau = 2$, the suspicion function for new unedited data is demonstrated by the dashed line. Tick marks denote the distances (median – lower quartile) and (upper quartile – median), respectively.*

SELEKT has an option to use (arithmetic) mean – standard deviation, mean + standard deviation and the mean instead of the quartiles and the median. A new version of SELEKT is planned that will make use of time-series analysis with forecasts.

Suspicion for the unedited value of a survey variable j to be in error can be assigned by more than one edit. SELEKT simply uses $\xi_{j,k,l} =$ the maximum of the suspicions $\xi_{i,k,l}$ produced by all edits associated to survey variable j, both traditional and the SELEKT-type edits. This is good enough in many cases, but not always. If a variable value in the current survey round grossly fails an edit check compared with the previous month but is very well accepted in a check with the same month last year, the received data can be accepted, that is, the minimum of the two computed suspicions would be appropriate. In cases like this, a composite edit rule of "traditional type" must be specified.

2.2.   Impact

This section begins by structuring the output statistics in Subsection 2.2.1. An erratic value in data can damage several estimated characteristics. In Subsection 2.2.2 the size of the error's impact on an estimate of a sum and on an estimate of a ratio are defined.

2.2.1.   Produced Statistics – Output

Statistical information is built up from structured sets of estimates of statistical characteristics, that is, statistical tables (Sundgren 2001). An estimated statistical characteristic is defined as a statistical measure applied to the values of one or more variables in a set of objects, a domain of study. The statistical measure is an aggregating function, for example a function that counts the number of units, a function that summarises the values of a variable, and a function that computes the ratio between the sums of two variables. The domains of study are constructed by variables that cross classify the population.

For SELEKT to be comfortably generic, an agreed organisation of the domains of study and functions is required. The domains must be clustered into *classes* in such a way that there is no overlap between the domains within a class, that is, no observed unit may contribute to more than one domain in a class. It is described in Subsection 2.3 that classes can be assigned a different importance that applies to all domains of each class.

Let d denote a domain in a class c of domains.

Example 5: Say that there are five domains in a survey, all based on Industry code; M = Manufacture, T = Trade, subdivided into W = Wholesale trade and R = Retail trade and finally All = M + T. A "typical" table looks like this:

Table 1. *Survey of employment and turnover (constructed data)*

Industry	Point estimates			Estimated standard errors		
	Sum of number of employees	Sum of turnover	Turnover per employee	Sum of number of employees	Sum of turnover	Turnover per employee
M	470	172,358	367	11.2	5,171	11.0
T = W + R	473	246,345	521	9.5	12,317	20.9
W	197	141,950	721	0.0	0	0.0
R	276	104,395	379	13.8	9,396	30.3
All = M + T	943	418,703	444	10.6	13,359	9.1

The five domains can be clustered into three classes in two alternative ways. One clustering, seemingly attractive, is {All}, {M, T} and {W, R}. The other way is {All}, {M, W, R}, {T} where the second class is composed of domains at various levels of detail. A class of domains need not cover all units, as exemplified by the class {W, R} which does not cover employees in Manufacture.

2.2.2. Impact on Estimated Statistical Characteristics

Impact measures how much the unedited instead of the edited value would affect the output statistics. Let

$y_{j,k,l}$ = the unedited value of variable j and unit k,l
$^e y_{j,k,l}$ = the edited value of variable j and unit k,l.

Estimate the sum of variable j in domain d in class c by summing the units in the sample that belong to the domain by:

$$\hat{T}_{c,d,j} = \sum w_{k,l} \cdot {}^e y_{j,k,l}$$

where $w_{k,l}$ is the combined sampling weight for the primary sampled unit k and observed unit l.

The impact on the estimate when retaining the unedited value for observed unit k,l belonging to c,d instead of the edited data value is

$$\eta_{c,d,j,k,l} = w_{k,l} \cdot \left(y_{j,k,l} - {}^e y_{j,k,l} \right)$$

In advance, $^e y_{j,k,l}$ is of course not available. As a proxy, an anticipated value $\tilde{y}_{j,k,l}^M$ is used, which can be the value from a previous survey round or an average of similar units in previous survey rounds. At this moment, it is not known whether the unedited value $y_{j,k,l}$ is in error or will be accepted. Potential impact is defined as:

$$^p \eta_{c,d,j,k,l} = w_{k,l} \cdot \left(y_{j,k,l} - \tilde{y}_{j,k,l}^M \right)$$

For the ratio of the sums of variables *j1* and *j2,* the potential impact of *k,l* is:

$$^p\eta_{c,d,j1,j2,k,l} = \frac{\hat{T}_{c,d,j1} + w_{k,l} \cdot \left(y_{j1,k,l} - \tilde{y}^M_{j1,k,l} \right)}{\hat{T}_{c,d,j2} + w_{k,l} \cdot \left(y_{j2,k,l} - \tilde{y}^M_{j2,k,l} \right)} - \frac{\hat{T}_{c,d,j1}}{\hat{T}_{c,d,j2}}$$

$$= w_{k,l} \cdot \frac{\left(y_{j1,k,l} - \tilde{y}^M_{j1,k,l} \right) - \frac{\hat{T}_{c,d,j1}}{\hat{T}_{c,d,j2}} \cdot \left(y_{j2,k,l} - \tilde{y}^M_{j2,k,l} \right)}{\hat{T}_{c,d,j2} + w_{k,l} \cdot \left(y_{j2,k,l} - \tilde{y}^M_{j2,k,l} \right)}$$

where $\hat{T}_{c,d,j1}$ and $\hat{T}_{c,d,j2}$ are estimated sums for variables *j1* and *j2* in domain *d* of class *c* from previous survey rounds. The ABS uses this potential impact as the local score according to Farwell (2004). A Taylor's approximation, which changes the denominator slightly if the anticipated error $y_{j2,k,l} - \tilde{y}^M_{j1,j2,g}$ is relatively small, is used on SELEKT:

$$^p\eta_{c,d,j1,j2,k,l} = w_{k,l} \cdot \frac{\left(y_{j1,k,l} - \tilde{y}^M_{j1,k,l} \right) - \frac{\hat{T}_{c,d,j1}}{\hat{T}_{c,d,j2}} \cdot \left(y_{j2,k,l} - \tilde{y}^M_{j2,k,l} \right)}{\hat{T}_{c,d,j2}}$$

If the anticipated error is large, the potential impact will be large anyway.

2.3. Local Scores

Early approaches to SE had two steps. First the units were either accepted or flagged for having at least one variable value in error. Second, a potential impact of each suspected unit on produced statistics was estimated. Only the units with the most potential impact on the statistics would be prioritised for manual follow-up. The ABS tool for significance editing, SigEE, works according to this approach (Farwell 2004).

SELEKT calculates an anticipated impact per variable, measure and domain as the product of suspicion and potential impact. The local score for the observed unit *k,l* and the estimated sum of variable *j* for the domain *c,d* to which unit *k,l* is assumed to contribute is now defined as:

$$SCORE5_{c,d,j,k,l} = \frac{\alpha_{c,d} \cdot \beta_j \cdot \left| \xi_{j,k,l} \cdot {}^p\eta_{c,d,j,k,l} \right|}{\max\left\{ SE(\hat{T}_{c,d,j}), \delta \cdot \hat{T}_{c,d,j} \right\}^{\gamma_j}}$$

where $SE(\hat{T}_{c,d,j})$ is the estimated standard error of the estimated sum of variable *j* in domain *c,d* from previous survey rounds, $\alpha_{c,d}$ is an importance weight for domain *c,d*, determined by subject-matter specialists and methodologists, β_j is a similar importance weight related to the estimated sum of variable *j*, also applied by Latouche and Berthelot (1992), δ and γ_j are parameters, $\delta > 0$ and $0 \leq \gamma_j \leq 1$.

This local score is the anticipated impact primarily relative to the standard error of the estimated sum. In sample surveys some domains may be totally enumerated, having zero standard errors. Some domains may have "accidentally" small estimated standard errors. SELEKT has one parameter δ to remedy these cases. A rather small value, like a general coefficient of variation in the survey, can be used as the value of δ. If sum is preferred to the standard error as the denominator, set δ to a big enough value, say 100.

SELEKT offers an option for assigning more importance to large domains and less importance to small domains, within the same class of domains. The denominator of the *SCORE*5 is raised to a power γ_j. The default value is 1. The value 0 reduces the denominator to a constant and anticipated impact is not related to any indicator of size of the domain. A "value in the middle" could be justified to obtain better relative quality of a large rather than a small domain, for example for "cars" rather than for "motorcycles" in the International Trade in Goods statistics.

In SE, as in survey design in general, it is necessary that the national statistical institute can assess the quality demands on each output table from the client's point of view.

The local score for two variables in a ratio estimated for a domain c,d is defined as

$$SCORE5_{c,d,j1,j2,k,l} = \frac{\alpha_{c,d} \cdot \beta_{j1,j2} \cdot \left| \xi_{j,k,l} \cdot^p \eta_{c,d,j1,j2,k,l} \right|}{\max \left\{ SE\left(\frac{\hat{T}_{c,d,j1}}{\hat{T}_{c,d,j2}} \right), \delta \cdot \frac{\hat{T}_{c,d,j1}}{\hat{T}_{c,d,j2}} \right\}^{\gamma_{j1,j2}}}$$

where $SE\left(\frac{\hat{T}_{c,d,j1}}{\hat{T}_{c,d,j2}} \right)$ is the estimated standard error of the estimated ratio and $\beta_{j1,j2}$ is the importance factor for the ratio of variables $j1$ and $j2$.

SELEKT produces a "weight matrix" with all the domains d as rows, grouped by classes c. Variables for which sums and ratios of sums are estimated are represented as columns. The cells of this weight matrix, denoted C, are $\alpha_{c,d} \cdot \beta_j / \max\{ SE(\hat{T}_{c,d,j}), \delta \cdot \hat{T}_{c,d,j} \}^{\gamma_j}$ for a sum and $\alpha_{c,d} \cdot \beta_{j1,j2} / \max\left\{ SE\left(\frac{\hat{T}_{c,d,j1}}{\hat{T}_{c,d,j2}} \right), \delta \cdot \frac{\hat{T}_{c,d,j1}}{\hat{T}_{c,d,j2}} \right\}^{\gamma_{j1,j2}}$ for a ratio.

The domain importance weights $\alpha_{c,d}$ are initially equal for all domains belonging to a class c. Once the matrix is calculated, the product manager is free to change any value in the matrix, for any reason.

Example 6 (cont. from Example 5): Number of employees and Turnover are measured for enterprises in a survey. Estimates are produced for five domains. The five domains are grouped into three classes in such a way that there is no overlap of domains within a class. Let us set the following parameter values: $\delta = 0.02$.

Table 2. *Importance weights for classes of domains*

Class of domains *(c)*	$\alpha_{c,d}$
c = 1: M, T	1
c = 2: W and R	1
c = 3: M + T	0.5

Table 3. *Importance weights for variables and measures*

	Variables/measures		
	Sum of Number of employees $j = 1$	Sum of Turnover $j = 2$	Turnover per employee $j1 = 2$ & $j2 = 1$
β_j, $\beta_{j1,j2}$	1	1	5
γ_j, $\gamma_{j1,j2}$	0.5	0.5	1

Now the resulting weight matrix C is computed (by SELEKT):

Table 4.   The weight matrix C (result)

Class and domain (c,d)	Sum of number of employees	Sum of turnover	Turnover per employee	Examples of computation
1, 1: M	0.299	0.0139	0.455	$1 \cdot 1/max\{11.2, 0.02 \cdot 470\}^{0.5} = 0.299$
1, 2: T	0.325	0.0090	0.240	$1 \cdot 1/max\{9.5, 0.02 \cdot 473\}^{0.5} = 0.325$
2, 3: W	0.504	0.0188	0.347	$1 \cdot 1/max\{0, 0.02 \cdot 197\}^{0.5} = 0.0188$
2, 4: R	0.269	0.0103	0.165	$1 \cdot 1/max\{9\,396, 0.02 \cdot 104\,395\}^{0.5}$ $= 0.0103$
3, 5: M + T	0.115	0.0043	0.278	$0.5 \cdot 5/ max\{9.1, 0.02 \cdot 444\}^{1} = 0.278$

When the matrix is computed by SELEKT, it is available and it is possible to alter any value, either ad hoc or by using an alternative model of computation.

2.4.   Aggregation of Local Scores to Global Scores

There is a hierarchy of scores with five levels. Let r denote respondent. In most surveys, the respondent answers for only one primary sampled unit, but there are examples where the respondent has many sampled units. The levels are:

   5: Domains (d),
   4: Variables/measures (j or $j1,j2$),
   3: Observed units (l)
   2: Primary sampled units (k)
   1: Respondent (r).

The numbering from 5 to 1 makes space for additional levels in future versions, if needed. Additional levels conceivably could be survey round and edit rule.

Scores for respondents, if being calculated, are global scores. Most often the scores for primary sampled units are considered the global scores. If the cost for recontacts in the follow-up workload is negligible, the scores for observed units can be used as global scores.

The local score for variable/measure j or $j1,j2$ is an aggregate of scores defined above:

$$SCORE4_{j,k,l} = \left\{ \sum_c \left[max\left(0, SCORE5_{c,d,j,k,l} - \theta_5\right) \right]^{\lambda_5} \right\}^{1/\lambda_5} \text{ or}$$

$$SCORE4_{j1,j2,k,l} = \left\{ \sum_c \left[max\left(0, SCORE5_{c,d,j1,j2,k,l} - \theta_5\right) \right]^{\lambda_5} \right\}^{1/\lambda_5}$$

where the summing $\sum_c$ is over all classes of domains (or domains since each selected unit contributes to one or no domain in each class of domains).

θ_5 is a threshold parameter. λ_5 is a parameter that defines the aggregation method. Hedlin (2008) presents the simple function used here to distinguish between three options. The value 1 implies sum, the value 2 implies sum of squares and the value 100 implies maximum (in practice). Latouche and Berthelot (1992) suggest the sum of local scores for variables, whereas Lawrence and McDavitt (1994) and Hedlin (2003) use the maximum of

the local scores. Farwell (2005) proposes the Euclidian distance, that is the value 2, as an alternative to the maximum when there are many scores at the variable level. SELEKT has four parameters; λ_5, λ_4, λ_3 and λ_2, one for each level to be aggregated.

The aggregation of local scores continues step by step up to the respondent level as:

$$SCORE3_{k,l} = \left\{ \sum_j \left[max\left(0, SCORE4_{j,k,l} - \theta_4\right) \right]^{\lambda_4} \right\}^{1/\lambda_4}$$

$$SCORE2_k = \left\{ \sum_l \left[max\left(0, SCORE3_{k,l} - \theta_3\right) \right]^{\lambda_3} \right\}^{1/\lambda_3}$$

$$SCORE1_r = \left\{ \sum_k \left[max\left(0, SCORE2_k - \theta_2\right) \right]^{\lambda_2} \right\}^{1/\lambda_2}$$

At all aggregations, SELEKT allows for a threshold value so that the maximum of zero and the score minus this threshold is aggregated. A value > 0 in early aggregation steps can be justified if there is a marginal cost of manual follow-up of extra variables, for example, with the constraint that the respondent has already been contacted. There are four threshold parameters θ_5, θ_4, θ_3 and θ_2. θ_1 is the threshold value for the global score, see Subsection 4.1.

$SCORE2_k$ is the global score in most surveys. An aggregation of scores for primary sampled units to $SCORE1_r$, where r primarily denotes respondent, is of interest in special cases. At Statistics Sweden the following situations have occurred:

- Respondents have several primary sampled units in the sample in the annual survey Rents for Dwellings. The national sample consists of approximately 12,000 rented dwellings. It turns out that there are only about 2,600 respondents, the real estate owners. The selective editing is focused on minimising the number of recontacts to respondents.
- The inflow of data for monthly surveys is irregular in the sense that two or more monthly forms can be delivered in the same "batch". Month is not defined as a level in the hierarchy of data. The $SCORE2_k$ would be an aggregate of scores for all months unless any action is taken. By defining Level 1 as the combination of primary sampled unit and month, that is a more detailed level than Level 2, the $SCORE1_r$ returns scores per primary selected unit and month.

3. Anticipated Value and a Measure of Variation

This section concerns methods and data to use for the computation of anticipated values and a basis for the SELEKT-type edits.

3.1. Time-Series Versus Cross-Section Data

There are quite different approaches to find anticipated values for the survey variables (j) and the test variables (i) and intervals of variation for the test variables for each observed unit s (k,l):

- Forecasts from the analysis of time-series data per observed unit.
- Latest observed values from previous survey rounds. Intervals of variation for the test variables cannot be computed.
- Cross-sectional analysis of data from previous survey rounds. Firstly find homogeneous groups of units, secondly compute the anticipated value as the median, the arithmetic mean or any other central value in the homogenous groups and also intervals of variation.

Intuitively, it makes sense to set the time-series measures as priority and the cross-sectional measures as reserve. Most business survey designs include updates of samples annually or more often. With a scheme of rotating samples, quite a significant proportion of units are new in the sample each year. This implies that different methods need to be used for different units if one wants to use the best method when possible.

3.2. Cold-Deck and Hot-Deck Data

The SE needs anticipated values and measures of variation before the editing of a survey round to make editing of data possible as soon as the first data arrive. Calculations of these measures are often based on edited data from past survey rounds, so-called cold-deck data. Generally calculations are made without using sampling weights. A decision has to be made whether to include imputed data. It seems most advisable from a theoretical point of view not to use imputed data, but it is easier not to distinguish between imputed and collected data. A decision must also be made about whether to include data that were suspicious but not flagged because the potential impact was low. Again it seems to be a good idea not to use highly suspected data, but it is easier not to make a difference.

For advanced time-series models, at least three years of monthly/quarterly data are needed. The simplest method in this context is to use the latest value or a simple function of it as anticipated value. If the survey measures phenomena with a heavy seasonal pattern, such as turnover in retail trade, a better alternative can be the value of the same month last year.

For cross-sectional data, one survey round can suffice, although several are preferable. Data from the current survey round, so-called hot-deck data, can be used by successively updating anticipated values and variation. The editing of prices for fresh fruits and vegetables in the consumer price index preferably makes use of robust means of current prices and price ratios. Prices from last month may be obsolete due to rapid price changes for fresh products.

3.3. Homogeneous Groups for Cross-Sectional Analysis

Cross-sectional analyses require homogeneous groups for which anticipated values are computed. The groups may, but need not, correspond to strata or domains of study. It is more important to stress homogeneity rather than to have a large number of observations in the homogeneous group. A well-known technique from the literature for imputation is to use the value of the very nearest neighbour as the anticipated value. However, by doing so, there is only one observation and it is not possible to compute an interval of variation for the SELEKT-type edits.

Homogeneous groups should be defined by multivariate analysis of cold-deck data. There are various methods that can be used. Norberg (2012) demonstrates regression tree analysis. The result of such an exercise is a new variable that identifies the groups.

SELEKT has one module that constructs joint homogeneous groups for the purposes of computing anticipated values and edit groups for SELEKT-type edits. Classificatory variables are listed by the user in hierarchical order and the minimum number of observations needed in the groups is specified. The cold-deck data are successively split by the classificatory variables, one by one, into one group for each value in the value set of the classificatory variable, as long as the condition for minimum number of observations is satisfied. The groups can be defined not only by different variables, but also by different numbers of digits within the classification codes.

3.4. Anticipated Value and Variation

For time-series data, there are a few options. The simplest is to use the edited values from the latest survey round for unit k,l. Here no measure of variation of the test variable can be computed. Another option is to use observations across a number of previous survey rounds and choose either the median, the lower and upper quartiles or arithmetic mean $\pm$ standard deviation. A third option is to produce a forecast for unit k,l by performing a time-series analysis including confidence intervals for the forecast.

For cross-sectional analysis, there are two natural sets of measures within the homogeneous groups; firstly, the lower quartile, median and upper quartile computed unweighted across the previous survey rounds for homogeneous groups, and secondly, the arithmetic mean and standard deviation.

SELEKT allows for auxiliary variable values $x_{j,k,l}$ which might help to compute the anticipated value $\tilde{y}_{j,k,l}^{M}$ of the $j{:}th$ variable. If the ratios of $y_{j,k,l}/x_{j,k,l}$ have a small variation, the anticipated value should firstly be found for the ratio and secondly this anticipated ratio should be multiplied by the individual $x_{j,k,l}$ to yield an anticipated value of $\tilde{y}_{j,k,l}^{M}$. The auxiliary variable should preferably be almost error free, so as not to cause many high scores by itself. The International Trade in Goods statistics are a converse example; the anticipated invoiced value of a transaction is the observed quantity multiplied by the median of price per quantity for transactions in the homogeneous group. There are more often errors in the quantity than in the invoiced value, but no better anticipated value can be found for invoiced value.

Anticipated values can be estimated by regression analysis with several explanatory variables. It is not a major problem to add files of anticipated values from tailor-made analysis to the SELEKT system.

As already noted in Subsection 3.1, one must be prepared to use different methods for computing the anticipated values and measures of variation. Different methods might be used for variables but particularly for various parts of the data. The SE can be inefficient if the "precision" of the anticipated values and intervals of variation varies. In Subsection 2.1.2 the indicator hit rate was defined. It is straightforward to compute this indicator for new units in the sample and those with a long series of data. Anticipated values, used in the impact dimension, can be analysed with edited values in scatter plots. The scores (being anticipated impacts relative to the standard error of the estimated sum) include both
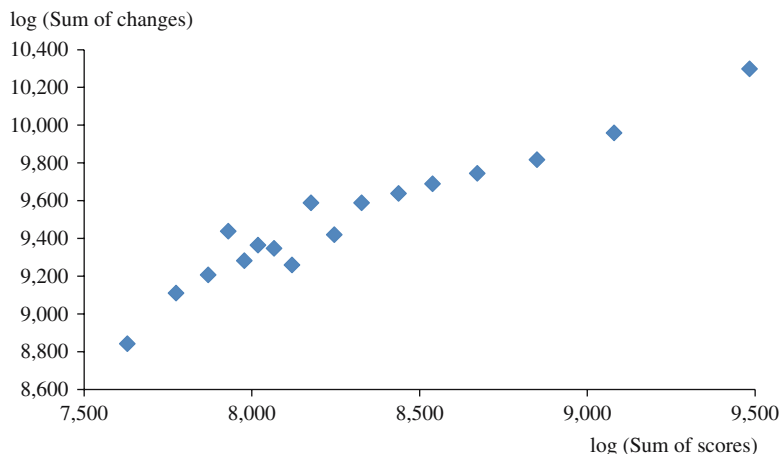
log (Sum of changes)

Fig. 3.   International Trade in Goods statistics. Vertical axis is sum of absolute values of changes on invoiced value. Horizontal axis is sum of scores. Each point represents an aggregate of 50 observed units for 70 months, in order of score. The very highest scores are far out and are not included in the graph.

dimensions. The graph in Figure 3 is a means to see if identified errors on average are proportional to scores for a survey with one prioritised variable, the sum of which is estimated. Some flagged units did not result in a change, some did. Batches of observations along the scale of score make computations of sum of impacts possible. As the scores in Figure 3 are fairly proportional to the changes on average, they could be considered useful predictors in the search for frequent and/or big errors. With this technique, it would also be possible to analyse any differences due to methods for estimation of anticipated values and so on by plotting two or more series of data. Surveys with many variables, many classes of domains and varying importance weights get a complex global score that it is scarcely possible to analyse this way.

4.   How Much is Enough?

This section presents two different ways to use the scores to identify data for follow-up. The traditional selective data editing approach selects units with scores above a threshold, called cut-off selection, as is discussed in Subsection 4.1. In Subsection 4.2 a new idea suggesting probabilistic editing is presented briefly. This second method is not yet implemented at Statistics Sweden.

4.1.   Cut-Off Selection

To determine cut-off thresholds for the global score, many calculations have to be carried out before the implementation using data from earlier survey rounds. Edited and unedited data are required. The concept of pseudobias is easy to understand and use.

Latouche and Berthelot (1992) define absolute pseudobias for an estimate when Q percent of the primary selected units with the highest scores are followed up as

$$\frac{\hat{T}_{c,d,j,Q} - \hat{T}_{c,d,j,Q=100}}{\hat{T}_{c,d,j,Q=100}}$$

where $\hat{T}_{c,d,j,Q}$ is an estimate of the sum of variable j in a domain c,d. $\hat{T}_{c,d,j,Q=100}$ is the estimated sum when all data that have been followed up are included (approximated by the old traditional heavy editing).

Lawrence and McDavitt (1994) define relative pseudobias, *RPB*, as

$$RPB_{d(s),j,Q} = \frac{\hat{T}_{c,d,j,Q} - \hat{T}_{c,d,j,Q=100}}{SE(\hat{T}_{c,d,j,Q=100})}$$

where $SE(\hat{T}_{c,d,j,Q=100})$ is the estimated standard error of $\hat{T}_{c,d,j,Q=100}$. Whenever $SE(\hat{T})$ in the denominator is close to zero, it should be replaced with some fraction of $\hat{T}$ as for SCORE5 in Subsection 2.3.

Särndal et al. (1992, 165) show that a 20% *RPB* has little effect on the coverage probability of an estimated confidence interval based on the estimated sampling variance. Allowing various importance parameters for (classes of) domains and variables/measures creates complications. It will not be as simple as requiring that all *RPBs* be less than 20%, but primarily the *RPBs* of the most important output should be less than 20%, while the rest can do with a somewhat higher *RPB*. Furthermore, the randomness of measurement errors in an evaluation data set occasionally causes high *RPB* in some domains. Evaluation as such should thus be done on more than one survey round. Hence, requiring *RPBs* less than 20% is a rule of thumb for important variables and domains, while accepting a few higher *RPBs* for less important statistical characteristics. Small domains can get a high *RPB* by accident even though errors are completely at random. When systematic errors frequently exist in data it will most likely be difficult to find a cut-off threshold that reduces the editing workload. MEMOBUST (2014) has three remarks on the method:

- The assumption of this approach is that the edited data can be considered 'true' data. This is a limitation because it rarely can be assumed.
- The simulation approach is frequently applied to data of a previous survey occasion to obtain a threshold value to be used for the current survey. It is worthwhile to note that in this case we assume that the error mechanism and the data distribution are the same on the two occasions.
- The method cannot be applied when you deal with the first wave of a survey.

For smaller recurrent surveys, such as industrial production, a feasible approach is to allow a real-time assessment by professional editing clerks based on the score itself, which may or may not be complemented by a ranking variable.

As a supplement to cut-off selection of units to follow up, a random sample beyond the threshold is useful for evaluating the performance of the SE in the long run. The results of the sample will indicate when the thresholds need to be updated with current data. Lewis (2014) discusses the need to formally maintain selective editing systems used in business surveys. When selective editing is first introduced, it offers the opportunity for efficient micro editing. However, without regular review, the thresholds can become out of date, potentially leading to an inefficient process and low-quality outputs.

4.2. *Probabilistic Editing*

Ilves and Laitila (2009) and Ilves (2010) propose quite a different editing procedure, where the responses are selected for editing through Poisson sampling according to their anticipated impact on final estimates, that is, the global scores. The probabilistic approach uses simple tools known from sampling theory to describe the effect of editing on the survey estimates. There is no need to restrict the follow-up to suspected data above a threshold. A two-phase design approach is applied to the bias estimation and a bias-corrected generalised regression (GREG) estimator and its variance are presented. Since the impact of systematic measurement errors is possible to estimate, it is not as important as for the cut-off method to identify and fix these errors before the selective editing.

5. Experience

Statistics Sweden has implemented selective editing in eleven surveys with extensive data editing over the last years and further surveys are in the pipeline for implementation (Norberg et al. 2014).

The experience is that implementation is a resource-intensive task that includes:

- staging tables of microdata and output statistics from SQL databases,
- creating performance indicators of the existing traditional edits,
- finding homogeneous edit groups by multivariate analysis,
- finding threshold values for local and global scores,
- integrating the SAS®-based SELEKT with production systems programmed in VB6 or VB.Net.

Efficient edit rules, as the result of reviewing the existing edits, are a basis for good data quality. The implementation of SE should at an early stage render in an evaluation of the existing traditional edit rules based on some hit-rate indicator. More efficient edit rules have been implemented as a result of these reviews.

It is necessary to address the question "Is selective editing appropriate for the survey?" as early as possible in the implementation stage. Statistics Sweden has developed a checklist based on experience. The checklist contains considerations of the following aspects:

- key variables are continuous,
- systematic measurement errors for known causes are dealt with in a first step,
- outputs are aggregates of microdata (statistical characteristics),
- anticipated values are possible to obtain.

The best profits were achieved for surveys where

- microediting is extensive, there is a potential for savings,
- the survey design is such that units have very different impacts on the estimates,
- there is a limited number of important variables and classes of domains, otherwise it will not be possible to define the parameters for the score function and global threshold so that the relative pseudobias is less than 20% percent for most statistical estimates.

The selective editing at Statistics Sweden has resulted in:

- reduction of error lists by 10–60% and consequently a reduction of cost for running production,
- a replacement of late macroediting by early microediting which implies that follow-ups can be done closer to data capture which is beneficent for respondents and measurement quality,
- more effective, more interesting and less stressful work for the editing staff,
- reduction of mishaps of records slipping through the editing since the editing staff now receive shorter error lists with a priority stated for all units.

6. References

Adolfsson, C. and P. Gidlund. 2008. "Conducted Case Studies at Statistics Sweden." Paper presented at the Work Session on Statistical Data Editing, Vienna, Austria, 21–23 April 2008. Available at: http://www.unece.org/fileadmin/DAM/stats/documents/2008/04/sde/wp.32.e.pdf (accessed February 2016).

Brinkley, E., K. Farwell, and F. Yu. 2011. "Selective Editing Methods and Tools: An Australian Bureau of Statistics Perspective." In Proceedings of Statistics Canada Symposium 2011. Available at: http://publications.gc.ca/collection_2013/statcan/11-522-x/CS11-522-2011-eng.pdf (accessed February 2016).

De Waal, T., J. Pannekoek, and S. Scholtus. 2011. *Handbook of Statistical Data Editing and Imputation*. New York: Wiley.

Di Zio, M. and U. Guarnera. 2013. "A Contamination Model for Selective Editing." *Journal of Official Statistics* 29: 539–555. Doi: http://dx.doi.org/10.2478/jos-2013-0039.

Farwell, K. and M. Raine. 2000. "Some Current Approaches to Editing in the ABS." In Proceedings ICES II of the Second International Conference on Establishment Surveys, Invited Papers. American Statistical Association 2001, 529–538.

Farwell, K. 2004. "The General Application of Significance Editing to Economic Collections." Australian Bureau of Statistics. Available at: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.260.1420 (accessed February 2016).

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 2005.

Granquist, L. 1984. "On the Role of Editing." *Statistical Review* 2: 105–118.

Granquist, L. 1997. "The New View on Editing." *International Statistical Review* 3: 381–387. Doi: http://dx.doi.org/10.2307/1403378.

Granquist, L. and J. Kovar. 1997. "Editing of Survey Data: How Much Is Enough?" *Survey Measurement and Process Quality*, 415–435. Doi: http://dx.doi.org/10.1002/9781118490013.ch18.

Hedlin, D. 2003. "Score Functions to Reduce Business Survey Editing at the UK Office for National Statistics." *Journal of Official Statistics* 19: 177–199.

Hedlin, D. 2008. "Local and Global Score Functions in Selective Editing." Paper presented at Work Session on Statistical Data Editing, Vienna, 21–23 April 2008.

Available at: http://www.unece.org/fileadmin/DAM/stats/documents/2008/04/sde/wp.31.e.pdf (accessed February 2016).

Hidiroglou, M.A. and J.-M. Berthelot. 1986. "Statistical Editing and Imputation for Periodic Business Surveys." *Survey Methodology* 12: 73–83.

Ilves, M. and T. Laitila. 2009. "Probability-Sampling Approach to Editing." *Austrian Journal of Statistics* 38: 171–182. Available at: http://www.stat.tugraz.at/AJS/ausg093/093Ilves.pdf (accessed February 2016).

Ilves, M. 2010. "Probabilistic Approach to Editing." Workshop on Survey Sampling Theory and Methodology, Vilnius, Lithuania, August 23–27, 2010. Available at: https://www.amstat.org/sections/srms/proceedings/y2010/Files/308253_60434.pdf (accessed February 2016).

Jäder, A. and A. Norberg. 2006. "A Selective Editing Method considering both Suspicion and Potential Impact, developed and applied to the Swedish Foreign Trade Statistics." Background facts on Economic Statistics 2006:3, Statistics Sweden. Available at: http://www.scb.se/statistik/_publikationer/OV9999_2006AOLBR_X100ST0603.pdf (accessed January 2016).

Latouche, M. and J.-M. Berthelot. 1992. "Use of a Score Function to Prioritize and Limit Re-Contacts in Business Surveys." *Journal of Official Statistics* 8: 389–400.

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

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

Lewis, D. 2014. "Maintenance of Selective Editing in ONS Business Surveys." Paper presented at Work Session on Statistical Data Editing, Paris, France, 28–30 April 2014. Available at: http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.44/2014/mtg1/Topic_1_ONS_Lewis.pdf (accessed February 2016).

MEMOBUST. 2014. *Handbook on Methodology of Modern Business Statistics CROS-portal*, Eurostat. Available at: https://ec.europa.eu/eurostat/cros/content/handbook-methodology-modern-business-statistics_eng (accessed February 2016).

Norberg, A. 2011. "The Edit." Paper presented at Work Session on Statistical Data Editing, Ljubljana, Slovenia, 9–11 May 2011. Available at: http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.44/2011/wp.26.e.pdf (accessed February 2016).

Norberg, A. 2012. "Tree Analysis – A Method for Constructing Edit Groups." Paper presented at Work Session on Statistical Data Editing, Oslo, Norway, 24–26 September 2012. Available at: http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.44/2012/05_Sweden.pdf (accessed February 2016).

Norberg, A., K. Lindgren, and C. Tongur. 2014. "Experiences from Selective Editing at Statistics Sweden." Paper presented at Work Session on Statistical Data Editing, Paris, France, 28–30 April 2014. Available at: http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.44/2014/mtg1/Topic_1_Sweden_Norberg.pdf (accessed February 2016).

Sundgren, B. 2001. "The $\alpha\beta\gamma\tau$-Model: A Theory of Multidimensional Structures of Statistics." Paper prepared for the MetaNet conference in Voorburg, the Netherlands, 2–4 April 2001. Available at: https://www.google.se/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwjb44jljfLKAhVjZ3IKHVUoC44QF

ggjMAA&url=https%3A%2F%2Fsites.google.com%2Fsite%2Fbosundgren%2Fmy-life%2FAlfaBetaGammaTauFinal.doc%3Fattredirects%3D0&usg=AFQjCNFKLgmtI-LR9YrLsQglsNWXnUY2mA&bvm=bv.114195076,d.bGQ (accessed February 2016).

Särndal, C.-E., B. Swensson, and J. Wretman. 1992. *Model Assisted Survey Sampling*. New York: Springer.

UNECE. 2000. *Glossary of Terms on Statistical Data Editing*. Available at: http://www.unece.org/fileadmin/DAM/stats/publications/editingglossary.pdf (accessed February 2016).