This module is part of the

Memobust Handbook

on Methodology of Modern Business Statistics

26 March 2014

Method: Minimum Adjustment Methods

Contents

General section

1. Summary

The problem of reconciling possibly conflicting information as described in the module "Micro-Fusion – Reconciling Conflicting Microdata" can be treated by an optimisation approach. In this approach, the values in the record with inconsistent microdata are changed, as little as possible, such that the modified record with microdata is consistent in the sense that it satisfies all edit rules. Formally then, the minimum adjustment method can be described as minimising a chosen distance between the original (inconsistent) record and the adjusted record, subject to the constraint that all edit rules are satisfied by the adjusted record. By specifying different distance functions, the minimum adjustment approach leads to different methods. For three common and, for the adjustment problem plausible, distance functions the corresponding adjustment methods will be described in this module and their differences will be illustrated by a numerical example.

2. General description of the method

2.1 Formal description of the optimisation problem

The optimisation approach resolves inconsistencies in data records with numerical variables that are required to adhere to a set of specified linear edit rules. The numerical variables in a record are denoted by x_i with $i = (1,...,n)$ and can be represented as a vector of variables: $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. The general form of a linear edit rule is as follows (see the module "Statistical Data Editing – Automatic Editing"):

$$e_{j1}x_1 + \cdots + e_{jn}x_n - c_j = 0\,, \tag{1}$$

for equalities and

$$e_{j1}x_1 + \cdots + e_{jn}x_n - c_j \geq 0 \tag{2}$$

for inequalities. Where $j = (1,...,J)$ numbers the edit rules, e_{ji} are numerical coefficients and c_j are numerical constants.

To describe the minimum adjustment methods it is convenient to express the edit rules in matrix notation. The equalities (1) can be expressed as $\mathbf{Ex} = \mathbf{c}$, with $\mathbf{E}$ the $J \times n$ "edit matrix" with elements e_{ji} and $\mathbf{c}$ the J-vector with elements c_j.

For the example record in table 1 of the module "Micro-Fusion – Reconciling Conflicting Microdata" we have

x = (1.*Profit,* 2.*Employees,* 3.*Turnover main,* 4.*Turnover other,* 5.*Turnover,* 6.*Wages,* 7.*Other costs,* 8.*Total costs*).

The three equality edits:

e_1: $x_1 - x_5 + x_8 = 0$ (*Profit = Turnover – Total Costs*)

e_2: $-x_3 + x_5 - x_4 = 0$ (*Turnover = Turnover main + Turnover other*)

e_3: $-x_6 - x_7 + x_8 = 0$ (*Total Costs = Wages + Other costs*)

can be expressed in the form **Ex = c** with

$$\mathbf{E} = \begin{pmatrix} 1 & 0 & 0 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & -1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & -1 & 1 \end{pmatrix} \text{ and } \mathbf{c} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Notice that the second column of **E** contains all zeroes because the second variable is not involved in any of the edit rules.

In this example a composite record was considered where three variables, *Turnover*, *Employees* and *Total costs* were obtained from reliable administrative sources and the other variables from a survey. As a consequence of obtaining the data from different sources, the edit rules are violated. The adjustment problem was to adjust the survey values such that the edit rules are satisfied while leaving the administrative values unchanged. For the optimisation approach it is necessary to take the distinction between free variables that are allowed to be adjusted and fixed variables that are not, into account. The complete data vector can be partitioned into $\mathbf{x}_{fre}$ for the free variables and $\mathbf{x}_{fix}$ for the fixed ones. A corresponding partitioning of the edit matrix yields, say, $\mathbf{E}_{fre}$ and $\mathbf{E}_{fix}$. Now we can write

$$\mathbf{Ex} = \mathbf{E}_{fre}\mathbf{x}_{fre} + \mathbf{E}_{fix}\mathbf{x}_{fix} = \mathbf{c},$$

and so $\mathbf{E}_{fre}\mathbf{x}_{fre} = \mathbf{c} - \mathbf{E}_{fix}\mathbf{x}_{fix}$,

*w*hich can be expressed as

$\mathbf{Ax}_{fre} = \mathbf{b}$, say.

The r.h.s. of this last expression contains all constants including the values of fixed variables and the l.h.s. contains the free variables that may be changed. They are the actual variables for the optimisation problem. For ease of notation we will, in the context of the optimisation problem, simply write **x** for the relevant, not fixed, variables and suppress the suffix *fre*. Thus we will write $\mathbf{Ax} = \mathbf{b}$ for the constraints on the relevant variables.

In addition to the equality constraints we also often have linear inequality constraints. The simplest case is the non-negativity of most economic variables. The optimisation approach can also handle linear inequality constraints. The constraints can then be formulated as $\mathbf{A}_{eq}\mathbf{x} = \mathbf{b}_{eq}$ and $\mathbf{A}_{ineq}\mathbf{x} \geq \mathbf{b}_{ineq}$, where $\mathbf{A}_{eq}$ contains the rows of **A** corresponding to the equality constraints and $\mathbf{A}_{ineq}$ the ones corresponding to the inequality constraints. For ease of exposition we shall, without noting otherwise, write these equality/inequality constraints more compactly as $\mathbf{Ax} \geq \mathbf{b}$

With the notation and conventions introduced above we can write the optimisation approach to the problem of finding the smallest possible adjustments compactly as

$$\begin{aligned} &\tilde{\mathbf{x}} = \arg\min_{\mathbf{x}} D(\mathbf{x}, \mathbf{x}_0) \\ &s.t. \;\; \mathbf{A}\tilde{\mathbf{x}} \geq \mathbf{b} \end{aligned}, \tag{3}$$

with $\mathbf{x}_0$ the adjustable part of the record *before* adjustment and $\tilde{\mathbf{x}}$ the corresponding sub-record *after* the adjustment and $D(\mathbf{x},\mathbf{x}_0)$ a function measuring the distance or deviance between $\mathbf{x}$ and $\mathbf{x}_0$. In the next section we will consider different functions D for the adjustment problem.

The conditions for a solution of the minimisation problem formulated in (3) can be found by inspection of the Lagrangian for this problem, which can be written as

$$L(\mathbf{x},\boldsymbol{\alpha}) = D(\mathbf{x},\mathbf{x}_0) + \boldsymbol{\alpha}^T(\mathbf{A}\mathbf{x}-\mathbf{b}), \tag{4}$$

with $\boldsymbol{\alpha}$ a vector of Lagrange multipliers, one for each of the constraints j.

From optimisation theory it is well known that for a convex function $D(\mathbf{x}, \mathbf{x}_0)$ and linear (in)equality constraints, the solution vector $\tilde{\mathbf{x}}$ must satisfy the so-called Karush-Kuhn-Tucker (KKT) conditions (see, e.g., Luenberger, 1984). One of these conditions is that the gradient of the Lagrangian w.r.t. $\mathbf{x}$ is zero when evaluated at the optimal point, i.e.,

$$L'_{x_i}(\tilde{x}_i,\boldsymbol{\alpha}) = D'_{x_i}(\tilde{x}_i,\mathbf{x}_0) + \sum\nolimits_j \alpha_j a_{ji} = 0, \tag{5}$$

with L'_{x_i} the gradient of L w.r.t. x_i and D'_{xi} the gradient of D w.r.t. x_i**.** From this condition alone, we can already see how different choices for D lead to different solutions to the adjustment problem. Below we shall consider three familiar choices for *D,* Least Squares, Weighted Least Squares and Kullback-Leibler divergence, and show how these different choices result in different structures of the adjustments, which we will refer to as the adjustment models. The form of these adjustment models gives some guidance to the choice of metric and the following properties may also be helpful in this respect. Weights in the WLS-criterion can be used to adjust some variables more than others, for instance because they are considered less reliable. Weights can also be used to make the amount of adjustment dependent on the size of the original value. Without knowledge about the preferred relative size of the adjustments for the different variables, the ordinary LS special case arises. The KL-criterion is only defined for positive variables: the original values need to be positive and the adjusted values are also guaranteed to be positive. The KL-adjustments can be expressed as positive multiplicative factors, larger original values will be adjusted more than smaller ones. More details of these adjustment models and their interpretation is given below.

2.2 Least squares adjustments

First, we consider the least squares criterion to find an adjusted x-vector that is closest to the original unadjusted data, that is: $D(\mathbf{x},\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}-\mathbf{x}_0)^T(\mathbf{x}-\mathbf{x}_0)$, and so $D'_{x_i}(\tilde{x}_i,\mathbf{x}_0) = \tilde{x}_i - x_{0,i}$, and we obtain from (5)

$$\tilde{x}_i = x_{0,i} + \sum\nolimits_j a_{ji}\alpha_j. \tag{6}$$

This shows that the least squares criterion results in an additive structure for the adjustments: the total adjustment to variable $x_{0,i}$ decomposes as a sum of adjustments to each of the constraints j. Each of these adjustments consists of an adjustment parameter α_j that describes the amount of adjustment due to constraint j and the entry a_{ji} of the constraint matrix **A** pertaining to variable i and constraint j Values of 1, –1 or 0 for a_{ji} imply that $x_{0,i}$ is adjusted by α_j, $-\alpha_j$ or not at all.

For variables that are part of the same constraints and have the same value a_{ji}, the adjustments are equal and the differences between adjusted variables are the same as in the unadjusted data. In particular, this is the case for variables that add up to a fixed total, given by a register value, and are not part of other constraints.

2.3 Weighted least squares adjustments

For the weighted least squares criterion, $D(\mathbf{x},\mathbf{x}_0)=\frac{1}{2}(\mathbf{x}-\mathbf{x}_0)^T Diag(\mathbf{w})(\mathbf{x}-\mathbf{x}_0)$, with $Diag(\mathbf{w})$ a diagonal matrix with a vector with weights along the diagonal. The derivative of this loss function in the optimum is $w_i(\tilde{x}_i - x_{0,i})$ and we obtain from (5)

$$\tilde{x}_i = x_{0,i} + \frac{1}{w_i}\sum_j a_{ji}\alpha_j . \tag{7}$$

Contrary to the least squares case where the amount of adjustment to a constraint is equal in absolute value (if it is not zero) for all variables in that constraint, the amount of adjustment now varies between variables according to the weights: variables with large weights are adjusted less than variables with small weights.

For variables that are part of the same constraints and have the same value a_{ji}, the adjustments are equal up to a factor $1/w_i$ and the differences of the weighted adjusted variables are the same as in the unadjusted data, that is, for variables i and i' we have $w_i\tilde{x}_i - w_{i'}\tilde{x}_{i'} = w_i x_{0,i} - w_{i'}x_{0,i'}$.

The weighted least squares approach to the adjustment problem has been applied by Thomson et al. (2005) in the context of adjusting records with inconsistencies caused by imputation. Some of the variables were missing and the missings were filled in by imputed values without taking care of edit constraints. This caused inconsistencies that were resolved by minimal adjustments, in principle to all variables, observed or imputed, according to the WLS-criterion. They used weights of 10,000 for observed values and weights of 1 for imputed values. Effectively, this means that if a consistent solution can be obtained by changing only imputed variables, this solution will be found. Otherwise (some of the) observed variables will also be adjusted.

One specific form of weights that is worth mentioning is obtained by setting the weight w_i equal to $1/x_{0,i}$ resulting, after dividing by $x_{0,i}$ in the adjustment model

$$\frac{\tilde{x}_i}{x_{0,i}} = 1 + \sum_j a_{ji}\alpha_j , \tag{8}$$

which is an additive model for the *ratio* between the adjusted and original values. It may be noticed that the expression on the right-hand side of (8) is the first-order Taylor expansion (i.e., around 0 for all the α_j's) to a multiplicative adjustment given by

$$\frac{\tilde{x}_i}{x_{o,i}} = \prod_j (1 + a_{ji}\alpha_j) \tag{9}$$

From (8) we see that the α_j 's determine the difference from 1 of the *ratio* between the adjusted and original values, which is usually much smaller than unity in absolute value (e.g., an effect of 0.2 implies a 20% increase due to adjustment which is large in practice). The products of the α_j 's are therefore often much smaller than the α_j 's themselves, in which cases (9) becomes a good approximation to (8), i.e., the corresponding WLS adjustment is roughly given as the product of the constraint-specific multiplicative adjustments.

2.4 Kullback-Leibler adjustments

The Kullback-Leibler divergence measures the difference between $\mathbf{x}$ and $\mathbf{x}_0$ by the function $D_{KL} = \sum_i x_i (\ln x_i - \ln x_{0,i} - 1)$. The derivative of this loss function is $\ln \tilde{x}_i - \ln x_{0,i}$ and we obtain from (5)

$$\tilde{x}_i = x_i \times \prod_j \exp\left(-a_{ji}\alpha_j\right). \tag{10}$$

In this case the adjustments have a multiplicative form and the adjustment for each variable is the product of adjustments to each of the constraints. The adjustment factor $\gamma_j = \exp(-a_{ji}\alpha_j)$ in this product represents the adjustment to constraint j and equals 1 if a_{ji} is 0 (no adjustment), $1/\gamma_j$ if a_{ji} is 1 and γ_k , if a_{ji} is -1.

For variables that are part of the same constraints and have the same value a_{ji}, the adjustments factors are equal and the ratios between adjusted variables are the same as between the unadjusted variables, $\tilde{x}_i / \tilde{x}_j = x_{0,i} / x_{0,j}$.

2.5 Generalisations: Adjusting to multiple sources and soft constraints

In this section we consider the possibilities for further modelling of the adjustment problem by using, simultaneously, information from multiple sources. First, we consider the situation that both register and survey values are considered to provide information for the final adjusted record rather than discarding survey values for which register values are available. Then we show that the approach used to combine information from multiple sources can be viewed as using, in addition to the "hard" constraints that are to be satisfied exactly, also "soft" constraints that only need to be fulfilled approximately.

2.5.1 Adjusting to both survey and register values

So far we considered the case where one of the sources (the administrative one) provides the reference values that are considered to be the correct ones and these values replace the values of the corresponding survey variables. Another situation arises when both data sources are considered to be fallible. In this situation we do not want to discard the data from one of the sources but we consider both sources to provide useful information on the variables of interest. This means that in the final consistent estimated vector we should not simply copy the values from the register values but obtain adjusted values that depend on both the survey values and the available register values. The data from the survey will be denoted by $\mathbf{x}_{0,S}$ and the data from the register by $\mathbf{x}_{0,R}$. In particular, for the example in table 1 of the module "Micro-Fusion – Reconciling Conflicting Microdata" we have the following:

$\mathbf{x}_{0,S}$ =(*Profit, Employees, Turnover main, Turnover other, Turnover, Wages, Other costs, Total costs*),
$\mathbf{x}_{0,R}$ =(*Employees_reg, Turnover_reg, Total costs_reg).*

where the suffix *_reg* is used to distinguish the register variables from their survey counterparts.

A consistent minimal adjustment procedure based on the information from both the survey values, the register values and the edit rules can be set up by considering the following constrained optimisation problem

$$\begin{aligned}&\tilde{\mathbf{x}} = \arg\min_{\mathbf{x}} \{D(\mathbf{x}, \mathbf{x}_{0,S}) + D(\mathbf{x}_R, \mathbf{x}_{0,R})\}\\ &\text{s.t. } \mathbf{Ax} \geq \mathbf{0}\end{aligned} \tag{11}$$

where the vector $\mathbf{x}_R$ denotes the subvector of $\mathbf{x}$ that contains the variables that are observed in the register. The vectors $\mathbf{x}$ and $\mathbf{x}_{0,S}$ both contain all variables and can be partitioned as $\mathbf{x} = (\mathbf{x}_{\bar{R}}^T, \mathbf{x}_R^T)^T$ and $\mathbf{x}_{0,S} = (\mathbf{x}_{0,S\bar{R}}^T, \mathbf{x}_{0,SR}^T)^T$, with $\bar{R}$ denoting the set of variables not in the register. Using this partitioning and the property that the distance functions considered in this paper are all decomposable in the sense that they can be written as a sum over variables, (11) can be re-expressed as

$$\begin{aligned}&\tilde{\mathbf{x}} = \arg\min_{\mathbf{x}} \{D(\mathbf{x}_{\bar{R}}, \mathbf{x}_{0,S\bar{R}}) + D(\mathbf{x}_R, \mathbf{x}_{0,SR}) + D(\mathbf{x}_R, \mathbf{x}_{0,R})\}\\ &\text{s.t. } \mathbf{Ax} \geq \mathbf{0}\end{aligned} \tag{12}$$

This clearly shows that the values of the variables *R* that are in both the register and the survey are adjusted to satisfy the edit constraints and remain as close as possible to both the register value and the survey value. Note that variables that are in both the register and the survey will be adjusted, if the two values are not equal, even if they do not appear in any edit rules, which is different from the situation considered before.

2.5.2 Soft constraints

The adjustment towards the register values due to a separate component in the objective function can also be interpreted as adding "soft" constraints to the optimisation problem. These soft constraints express that $\tilde{\mathbf{x}}_R$ should be approximately equal to the register values $\mathbf{x}_{0,R}$ but need not "fit" these data exactly as was required before.

The notion of soft constraints opens up a number of possibilities for further modelling the adjustment problem. Suppose, for instance, that the total amount of wages paid (*Wages*) is known from an administrative source and treated as fixed while the number of employees (*Employees)* is a free variable. Furthermore, assume that before adjustment the wages are 20,000 Euros per employee and that it is plausible that this ratio should hold approximately for the record after adjustment. This can be formulated as a "soft" ratio constraint on *Employment* and *Wages*: *Wages* / *Employment* ≈ 20,000. This soft constraint can be handled by the optimisation problem by adding to the loss function the component $D(x_{wages}, 20000 \times x_{employment})$. This soft constraint is often more reasonable than using hard upper and lower bounds on the adjusted value for *Employment*. In fact we can do both, for instance to bound *Employment* within certain hard limits and use the soft constraint to draw the value of *Wages* within these bound towards the expected value of 20,000 times the number of employees.

3. Preparatory phase

4. Examples – not tool specific

4.1 Comparison of distance functions using the example record

The different methods (LS, WLS and KL) have been applied to make the two composite records consistent that are in the example of table 1 in the module "Micro-Fusion – Reconciling Conflicting Microdata". For the WLS method we used as weights the inverse of the $\mathbf{x}_0$-values so that the relative differences between $\mathbf{x}$ and $\mathbf{x}_0$ are minimised and the adjustments are proportional to the size of the $\mathbf{x}_0$-values.

The optimisation methods were implemented by an iterative method which is a special case of the so-called row-action algorithms treated in Censor and Zenios (1997) (see also, De Waal et al., 2011, Ch. 10). For the (weighted) least squares adjustments an R-package is available (van der Loo, 2012).

The results for the different methods are in table 1 below. The solutions for the KL- and WLS-adjustments appeared to be the same in all digits shown and were therefore combined into a single column. With the weights used here these solutions should be similar in practice. The register values that are treated as fixed are shown in bold; the other values may be changed by the adjustment procedure.

Table 1. Example business record: two composite versions and adjusted values.

Variable	*Name*	*Composite record II*			*Composite record I*		
		Unadj.	LS	WLS/KL	Unadj.	LS	WLS/KL
x_1	Profit	330	282	291	330	260	249
x_2	Employees	20	20	20	**25**	**25**	**25**
x_3	Turnover main	1000	960	922	1000	960	922
x_4	Turnover other	30	-10	28	30	-10	28
x_5	Turnover	**950**	**950**	**950**	**950**	**950**	**950**
x_6	Wages	500	484	470	**550**	**550**	**550**
x_7	Other costs	200	184	188	200	140	151
x_8	Total costs	700	668	658	700	690	701

Unadj. = Unadjusted values.
LS = adjusted values according to the LS criterion.
WLS/KL = adjusted values according to the WLS or KL criterion.

For both composite records, the LS adjustment procedure leads to one negative value for *Turnover other*, which is not allowed for this variable. Therefore the LS-procedure was run again with a non-negativity constraint added for the variable *Turnover other.* This results simply in a zero for that variable and a change in *Turnover main* to ensure that *Turnover* = *Turnover main* + *Turnover other.* Without the non-negativity constraint, the LS-results clearly show that for variables that are part of the same constraints (in this case the pairs of variables x_3, x_4 and x_6, x_7 that are both appearing in one constraint only), the adjustments are equal: –40 for x_3, x_4 and –16 for x_6, x_7. *Total costs* (x_8) is part of two constraints and therefore the total adjustment to this variable consists of two additive components. One component to adjust to the constraint e_1: $x_1 - x_5 + x_8 = 0$ (*Profit = Turnover – Total Costs*) and one component to adjust to e_3: $x_8 - x_6 - x_7 = 0$ (*Total Costs = Wages + Other costs*). For the composite record II, the first component is minus 48 – which is also the single adjustment component for *Profit* – and the second component is 16 – which is also the single adjustment component for *Wages* and *Other costs* (with opposite sign). These two components add up to the adjustment of –32.

The results for the WLS/KL solution show that for this weighting scheme the adjustments are larger, in absolute value, for large values of the survey variables than for smaller ones. In particular, the adjustment to *Turnover other* is only –2.3 – so that no negative adjusted value results in this case – whereas the adjustment to *Turnover main* is 77.7. The multiplicative nature of these adjustments (as KL-type adjustments) also clearly shows since the adjustment *factor* for both these variables is 0.92 (for both composite records). The adjustment factor for *Wages* and *Other costs* in composite record I is also equal (to 0.94) because these variables are in the same single constraint and so the ratio between these variables is unaffected by this adjustment. However the ratio of each of these variables to *Total Costs* is not unaffected because *Total Costs* has a different sign in the constraint e_3 and, moreover, *Total Costs* is also part of constraint e_1 so that it is subject to two adjustment factors.

5. Examples – tool specific

6. Glossary

For definitions of terms used in this module, please refer to the separate "Glossary" provided as part of the handbook.

7. References

Censor, Y. and Zenios, S. A. (1997), *Parallel Optimization. Theory, Algorithms, and Applications*. Oxford University Press, New York.

De Waal, T., Pannekoek, J., and Scholtus, S. (2011), *Handbook of Statistical Data Editing and Imputation*. John Wiley & Sons Inc., Hoboken, New Jersey.

Luenberger, D. G. (1984), *Linear and Nonlinear programming, second edition*. Addison-Wesley, Reading.

Pannekoek, J. (2011), Models and algorithms for micro-integration. In: *Report on WP2: Methodological developments*, ESSNET on Data Integration, available at http://www.cros-portal.eu/content/wp2-development-methods.

Pannekoek, J. and Zhang, L.-C. (2011), Partial (donor) imputation with adjustments. Working Paper No. 40, UN/ECE Work Session on Statistical Data Editing.

van der Loo, M. (2012), *rspa: Adapt numerical records to fit (in)equality restrictions with the Successive Projection Algorithm*. R package version 0.1-1.

Thomson, K., Fagan, J. T., Yarbrough, B. L., and Hambric, D. L. (2005), Using a Quadratic Programming Approach to Solve Simultaneous Ratio and Balance Edit Problems. Working paper 32, UN/ECE Work Session on Statistical Data Editing, Ottawa.

Specific section

8. Purpose of the method

The purpose of the method is to adjust the values of some variables in a data record to remove edit violations to ensure consistency of the data values obtained from different sources.

9. Recommended use of the method

1. The method should be used after detection and treatment of errors and missing values.

10. Possible disadvantages of the method

1. When inconsistencies arise due to large errors in some values, these errors may propagate to other values due to adjustment. Influential errors should therefore be treated before the method is applied.

11. Variants of the method

1. Least squares adjustments
2. Weighted least squares adjustments.
3. Kullback-Leibler adjustments.

12. Input data

1. Data records with possibly inconsistent values and edit rules.

13. Logical preconditions

1. Missing values
    1. Missing values are allowed but edit rules involving variables with missing values cannot be checked and no adjustment with respect to these edit rules will take place.
2. Erroneous values
    1. Influential erroneous values should be treated before the method is applied.
3. Other quality related preconditions
    1.
4. Other types of preconditions
    1.

14. Tuning parameters

1. The amount of change applied to individual variables can be controlled by specifying weights for the variables

15. Recommended use of the individual variants of the method

1.

16. Output data

1. The output consists of the same individual records as the input, with values adapted when needed to ensure consistency with the edit rules.

17. Properties of the output data

1. The output data are ensured to be consistent with all specified edit rules that do not involve variables with missing values.

18. Unit of input data suitable for the method

The input consists of individual records that are treated one-by-one, independently.

19. User interaction - not tool specific

1.

20. Logging indicators

1.

21. Quality indicators of the output data

1.

22. Actual use of the method

1.

Interconnections with other modules

23. Themes that refer explicitly to this module

1. Micro-Fusion – Data Fusion at Micro Level
2. Statistical Data Editing – Main Module
3. Statistical Data Editing – Automatic Editing
4. Statistical Data Editing – Editing Administrative Data
5. Imputation – Main Module

24. Related methods described in other modules

1. Micro-Fusion – Reconciling Conflicting Microdata
2. Micro-Fusion – Prorating
3. Micro-Fusion – Generalised Ratio Adjustments

25. Mathematical techniques used by the method described in this module

1. Optimisation of convex functions with linear (in)equality constraints.

26. GSBPM phases where the method described in this module is used

1. Phase 5 - Process

27. Tools that implement the method described in this module

1. The R-package rspa of van der Loo (2012) can be used to apply adjustment according to the (weighted) least squares criterion.

28. Process step performed by the method

GSBPM Sub-process 5.3: Review, validate and edit

Administrative section

29. Module code

Micro-Fusion-M-Minimum Adjustment Methods

30. Version history

Version	Date	Description of changes	Author	Institute
0.1	05-03-2013	first version	Jeroen Pannekoek	CBS (Netherlands)
0.2	17-04-2013	second version	Jeroen Pannekoek	CBS (Netherlands)
0.2.1	09-09-2013	preliminary release		
0.3	20-12-2013	improvements based on the EB-review	Jeroen Pannekoek	CBS (Netherlands)
1.0	26-03-2014	final version within the Memobust project		

31. Template version and print date

Template version used	1.0 p 4 d.d. 22-11-2012
Print date	21-3-2014 18:01