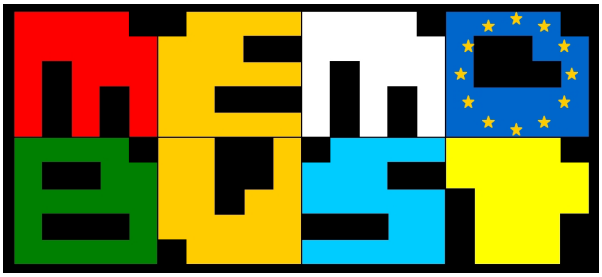

This module is part of the

Memobust Handbook

on Methodology of Modern Business Statistics

26 March 2014

Method: Generalised Ratio Adjustments

Contents

General section 3

1. Summary 3
2. General description of the method 3
3. Preparatory phase 5
4. Examples – not tool specific 5

4.1 Generalised ratio adjustment compared with WLS/KL-adjustments 5

5. Examples – tool specific 6
6. Glossary 6
7. References 6

Specific section 7

Interconnections with other modules 8

Administrative section 10

General section

1. Summary

Generalised ratio adjustment is a method to reconcile conflicting information as described in the module "Micro-Fusion – Reconciling Conflicting Microdata". The method uses multiplicative adjustments, just as the methods prorating (see the module "Micro-Fusion – Prorating") and the KL-adjustments (see the module "Micro-Fusion – Minimum Adjustment Methods"). The generalised ratio adjustments method aims to make the adjustments as uniform as possible. Furthermore, and in contrast with the other adjustment methods, the method can result in adjustments to variables that are not involved in any of the constraints.

2. General description of the method

The generalised ratio adjustments are multiplicative adjustments applied to variables that are "free" variables which means that they are designated to be adjustable. These variables may or may not be involved in edit constraints. The adjustments methods considered in the modules "Micro-Fusion – Prorating" and "Micro-Fusion – Minimum Adjustment Methods" were only meant to resolve violations of edit rules, therefore only free variables involved in edit rules were adjusted since variables not appearing in the edit rules are irrelevant because they cannot violate edit rules.

However, there may be reasons other than the violation of edit rules to change the values of some variables. Consider, for instance, the business record shown in section 4 below (and in table 1 of the module "Micro-Fusion – Minimum Adjustment Methods"). In the column denoted by *Survey*, values for the variables are shown that are obtained from a survey. Two scenarios are assumed for additional data: (I) from administrative sources, values are available for the variables *Employees, Turnover* and *Wages* (the values in bold in the columns *Adjusted Composite (I)*) and (II) an administrative source is only available for the variable *Turnover*. Suppose that the administrative data are treated as fixed, for instance because they are more recent (although less detailed) and / or more accurate than the survey data. Adjusting the values of the survey variables *Turnover main* and *Turnover other* can then be seen as extrapolating the (slightly) outdated survey values to the more recent administrative data. Apparently, according to the available data, this unit's turnover has been reduced (from 1030 to 950) and multiplicative adjustments for *Turnover main* and *Turnover other* are easily obtained by reducing them by the same ratio of 1030/950. In this case one may be tempted to apply this rescaling to all variables, also those not involved in constraints, which can be justifiable if it is assumed that these variables are related to *Turnover* in approximately the same way as in the original survey record; in some sense the "size" of the business has decreased by a factor 1030/950 and all variables are scaled with this factor to reflect this change. In the newly created consistent record their ratio to *Turnover* would be preserved by this rescaling. This intuitive and simple solution becomes difficult if more variables are obtained from administrative sources leading to multiple adjustment factors. It is then not obvious how to take these different factors into account and to ensure that constraints are satisfied.

One possible solution is to use the minimum adjustment approach described in the module "Micro-Fusion – Minimum Adjustment Methods" and to add the ratios of the variables not involved in the constraints to each of the administrative variables as "soft" constraints to the optimisation problem (see section 2.5.2 of that module). However, this approach leads to a non-trivial modelling effort and a more complicated loss function. As a method that can be applied more routinely, using only the

already specified edit constraints, Pannekoek and Zhang (2011) suggested a generalised ratio adjustments method. As in the modules "Micro-Fusion – Reconciling Conflicting Microdata" and "Micro-Fusion – Minimum Adjustment Methods" a composite record is considered, consisting of values obtained from different sources, that may violate some linear edit constraints. The task is to make adjustments to a subset of the variables in the composite record such that the resulting record becomes consistent with the edit rules. The variables that are allowed to be adjusted are named *free* variables and the other variables are the *fixed* variables. For instance in scenario *(I)* the unadjusted composite record consists of values from the administrative source for variables x_2, x_5 and x_6 and these variables are treated as fixed. The remaining variables in the composite record have values from the survey; these variables are treated as free and will be adjusted to meet the edit constraints. The generalised ratio adjustments method finds multiplicative adjustments such that the resulting adjusted values meet the following two requirements: (1) the edit-constraints are satisfied and (2) the changes with respect to the original survey record are as uniform as possible (resembling a uniform overall ratio adjustment as much as possible).

The generalised ratio adjustments method focusses on the changes between the values in the *original survey record* and the final *adjusted composite record*. These changes can be expressed as factors δ_i, defined by

$$\delta_i = \tilde{x}_i / x_{s,i} \text{ , for } i = 1, \cdots, n \text{ ,} \qquad (1)$$

with n the number of variables, $\tilde{x}_i$ the values of the variables in the adjusted composite record and $x_{s,i}$ the survey values for these variables. By definition, the values of the fixed variables in the composite records are the same before and after adjustment. For these variables, the change factors δ_i represent the change between the survey value and the administrative value. For free variables the changes δ_i are adjustment factors that adjust the survey values such that the edit constraints are satisfied.

Before adjustment, the composite record consists of values $x_{0,i}$ which are equal to the administrative values if these are available and equal to the survey values otherwise. The record x_0 differs from the original survey record in the administrative values only. Since the administrative values are treated as fixed, these values will not be changed by the adjustment procedure and thus the change factors for the fixed variables can be expressed as

$$\delta_i = \tilde{x}_i / x_{s,i} = x_{0,i} / x_{s,i} \text{ , for } i \in I_{fix} \text{ ,} \qquad (2)$$

with I_{fix} the set of indices corresponding to the fixed (administrative) variables. For the fixed variables the δ_i are given by (2) but for the other variables, the free variables with index set I_{free}, the δ_i need to be determined such that the edit rules are satisfied and all change factors (including those for the fixed variables) are as uniform as possible. Specifically, the δ_i will be obtained by minimising the following objective function (Δ) over the δ_i corresponding to the free variables:

$$\min_{\delta_i | i \in I_{free}} \Delta = \sum_{i \in I_{free}} (\delta_i - \bar{\delta})^2 \text{ , where } \bar{\delta} = \frac{1}{n} \sum_{i=1}^{n} \delta_i \text{ ,} \quad (3)$$

with constraints as in the module "Micro-Fusion – Minimum Adjustment Methods", i.e., $\mathbf{A}\tilde{\mathbf{x}}_{free} = \mathbf{b}$ with $\tilde{\mathbf{x}}_{free}$ the vector with adjusted free variables. Notice that the minimum is taken over the free variables only but the mean is taken over all variables, both free and fixed. The objective function (3) can be viewed as a function of the change factors δ_i for the free variables but also as a function of the adjusted values $\tilde{x}_i$ (since $\delta_i = \tilde{x}_i / x_{s,i}$) for these variables. In either case the variation in the change factors is minimised subject to the linear edit constraints on the adjusted values. A possible generalisation of (3) that differentiates between the effects of the different δ_i on the objective value is to use weights similar to the WLS loss-function in "Micro-Fusion – Minimum Adjustment Methods".

The fact that the objective function makes the changes (with respect to the original survey record) for all variables in the record as uniform as possible results in two properties of the generalised ratio adjustments not shared by the minimum adjustment methods. Firstly, adjustments are defined for all free variables, whether they are involved in edit constraints or not. This is because minimising the variation in the δ_i will, in general, lead to values for δ_i unequal to 1 (and hence to adjustment) even for survey values not involved in edit constraints. Secondly, the information from the changes between the survey values and administrative values of the fixed variables is used in the adjustment procedure. This is because the mean of all changes, $\bar{\delta}$, is partly determined by these changes in the fixed variables and therefore these changes influence the adjustment factors for the free variables since they are made to vary as little as possible around $\bar{\delta}$.

3. Preparatory phase

4. Examples – not tool specific

4.1 Generalised ratio adjustment compared with WLS/KL-adjustments

In this example we show the results of the generalised ratio method and compare these results with the WLS/KL-adjustments described in the module "Micro-Fusion – Minimum Adjustment Methods". Both methods use multiplicative adjustments but the WLS/KL-adjustments apply only to variables that are involved in constraints whereas the generalised ratio method can also adjust variables that are not involved in constraints and, in addition, this last method will result in adjustments that are as uniform as possible. Both methods will result in a record that satisfies all linear constraints.

The data for this example are the values of a business record shown in table 1 of module "Micro-Fusion – Reconciling Conflicting Microdata" and repeated in Table 1 below. Two versions of an adjusted composite record are shown[1], one for a record with three values obtained from an administrative source (which are shown in in bold) that is denoted by *Adjusted Composite (I)* and another with only *Turnover* obtained from an administrative source, denoted by *Adjusted Composite*

[1] Values are rounded to the nearest integer.

(II). The other values are from a survey, see the column *Survey*. The administrative values are treated as fixed while the survey values are free, i.e., they can be adjusted.

The composite record *(II)* with only *Turnover* from the administrative source violates two edit rules:

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*);

The survey value of *Turnover* is 1030 and, as expected, the generalised ratio adjustments for this record reduce to a global proportional adjustment of all the survey values by a ratio of 0.922 (=950/1030) including the variable *Employee*. That this last variable is adjusted is a difference with the minimum-adjustment approaches that only adjust variables that are involved in constraints.

Table 1. Example business record with survey values and adjusted values for the WLS/KL and generalised ratio methods.

Variable	*Name*	*Survey*	*Adjusted Composite (I)*		*Adjusted Composite (II)*	
			WLS/KL	Gen. Ratio	WLS/KL	Gen. Ratio
x_1	Profit	330	249	239	291	304
x_2	Employees	20	**25**	**25**	20	18
x_3	Turnover main	1000	922	921	922	922
x_4	Turnover other	30	28	29	28	28
x_5	Turnover	1030	**950**	**950**	**950**	**950**
x_6	Wages	500	**550**	**550**	470	461
x_7	Other costs	200	151	161	188	184
x_8	Total costs	700	701	711	658	646

For composite record *(I)* with *Turnover*, *Wages* and *Employees* obtained from administrative sources, three edit rules are violated: in addition to e_1 and e_2 also the rule

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

is violated. Also in this case, the generalised ratio adjustments are close to the WLS/KL solution. The empirical variance of the multiplicative factors (i.e., proportional to the value of the loss function Δ) is 0.0270 for the generalised ratio adjustments, which is a little bit less than the value 0.0276 obtained for the WLS/KL solution.

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

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

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. The generalised ratio adjustments method aims to make the adjustments as uniform as possible. Furthermore, and in contrary to the other adjustment methods, the method can result in adjustments to variables that are not involved in the constraints.

9. Recommended use of the method

1.

10. Possible disadvantages of the method

1.

11. Variants of the method

1.

12. Input data

1.

13. Logical preconditions

1. Missing values
    1.
2. Erroneous values
    1.
3. Other quality related preconditions
    1.
4. Other types of preconditions
    1.

14. Tuning parameters

1.

15. Recommended use of the individual variants of the method

1.

16. Output data

1.

17. Properties of the output data

1.

18. Unit of input data suitable for the method

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 – Editing Administrative Data
4. Imputation – Main Module

24. Related methods described in other modules

1. Micro-Fusion – Reconciling Conflicting Microdata
2. Micro-Fusion – Prorating
3. Micro-Fusion – Minimum Adjustment Methods

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

1. Quadratic optimisation

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. There are no specific tools available that implement this method. However, the method can be applied using quadratic programming routines.

28. Process step performed by the method

GSBPM Sub-process 5.3: Review, validate and edit

Administrative section

29. Module code

Micro-Fusion-M-Generalised Ratio Adjustments

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)
0.3	09-07-2013	third version	Jeroen Pannekoek	CBS (Netherlands)
0.3.1	09-09-2013	preliminary release		
0.4	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