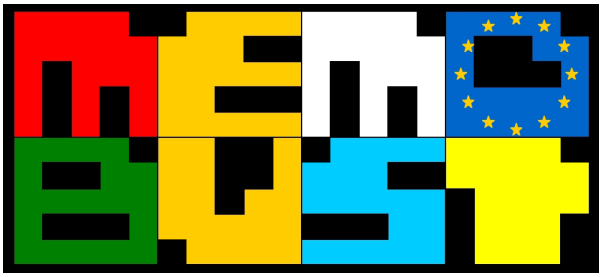

This module is part of the

Memobust Handbook

on Methodology of Modern Business Statistics

26 March 2014

Method: Prorating

Contents

General section

1. Summary

Prorating is a simple method to reconcile conflicting information as described in the module "Micro-Fusion – Reconciling Conflicting Microdata". The method is designed for equality edits, especially with business statistics in mind, where often a total (turnover, costs etc.) is broken down into a number of specifications (turnover from different activities, different kinds of costs). Inconsistencies arising when the specifications do not add up to the total are often handled by prorating. The method handles a single edit rule at a time and is therefore in practice applied to each of the edit rules one by one. This has the drawback that the order in which the edits are treated does matter and quite different results can be obtained by different orders. This drawback has led to the more principled approaches described in the modules "Micro-Fusion – Minimum Adjustment Methods" and "Micro-Fusion – Generalised Ratio Adjustments".

2. General description of the method

2.1 The prorating method

Consider the following situation described in the module "Micro-Fusion – Reconciling Conflicting Microdata". In a business data set obtained by linking a survey to an administrative source, we observe for some unit the following values for three variables (x_3, x_4, x_5) describing turnover:

x_3: *Turnover main*	x_4: *Turnover other*	x_5: *Turnover total*
1000	30	950

The variable *Turnover total* is obtained from an administrative source while the component variables are observed in a survey. An inconsistency arises because the sum of x_3 and x_4 is 1030 instead of 950 and the edit rule $x_5 = x_3 + x_4$ is violated. Suppose that the administrative value of *Turnover total* is not to be changed but the other values may be changed in order to make the record consistent. The prorating method (Banff Support Team, 2008; Pannekoek, 2011; Pannekoek and Zhang, 2011) changes the adjustable values by a uniform multiplicative adjustment. Thus, in this case, the adjusted values for x_3 and x_4 become (950/1030)×1000 and (950/1030)×30.

For a general description it is convenient to express the equality edit in the form $\sum_i x_i = 0$, which involves changing the sign of some of the original variables. In the example above this could be accomplished by defining $x_5 + (-x_3) + (-x_4) = 0$. Furthermore, let δ denote the prorating factor and let I_{fre} and I_{fix} be the index sets of, respectively, the adjustable (free) variables and the un-adjustable (fixed) variables. Then, the adjusted values are given by

$$\tilde{x}_i = \delta x_i \text{ for } i \in I_{fre}. \qquad (1)$$

Now, since we must have $\sum_{i \in I_{fre}} \tilde{x}_i + \sum_{i \in I_{fix}} x_i = 0$ for the adjusted values to satisfy the equality edit, we can write:

$\delta \sum_{i \in I_{fre}} x_i = -\sum_{i \in I_{fix}} x_i$, and so,

$$\delta = -\sum_{i \in I_{fix}} x_i \,/\, \sum_{i \in I_{fre}} x_i \,. \tag{2}$$

From (1) we can see that for a solution to this adjustment problem it is necessary that there are free variables with non-zero values which is understandable because a multiplicative adjustment would otherwise be ineffective.

2.2 Weighted prorating

A weighted version of the prorating method makes it possible to control the relative amount of change in the free variables. A weight is assigned to each free variable and the amount of change is inversely proportional to the weight.

In this case we can write, for the adjusted values,

$$\tilde{x}_i = \frac{\delta}{w_i} x_i \text{ for } i \in I_{fre} \,. \tag{3}$$

Furthermore, since we must have $\sum_{i \in I_{fre}} \tilde{x}_i = -\sum_{i \in I_{fix}} x_i$, we obtain the following expression for δ:

$$\delta = -\sum_{i \in I_{fix}} x_i \,/\, \sum_{i \in I_{fre}} \frac{x_i}{w_i} \,. \tag{4}$$

3. Preparatory phase

4. Examples – not tool specific

4.1 Prorating applied in two different orders

Prorating is defined as a treatment for a single edit inconsistency. It also applies to several edit inconsistencies without complications as long as the edits have no variables in common. However, it does not, in itself, provide a unique solution for systems of connected edits. For such cases, a strategy is followed that involves treating the edits in a predefined order and fixing each variable that has been treated (see Banff Support Team, 2008). This is illustrated in the example below.

In this example we show the results of applying prorating with two different orders to resolve the violation of the edit rules for the values of the business record shown in Table 1 of the module "Micro-Fusion – Reconciling Conflicting Microdata", column *Composite (I)*. The data in this column consist of administrative values for the variables in bold in Table 1 below, *Employees*, *Turnover* and *Wages*, and values observed in a survey for the other variables. This composite record violates three 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*);

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

Now, we assume that the administrative values are fixed and adjust the other values by prorating so that the three edit rules are satisfied. The result for the edit e_2 is independent of the order in which prorating is applied because the free variables in this edit do not appear in other edits and are only adjusted to sum up to the total *Turnover.* The order in which the edit rules e_1 and e_3 are treated does make a difference for the result because these variables have a free variable (*Total costs*) in common. If a top-down strategy is followed in which first the edit e_1 is treated (which entails adjustment of *Profit* and *Total costs*) and then the edit e_3 is treated (which amounts to adjustment of *Other costs*), we obtain the results in the column "e_1 *adjusted first*". If the prorating adjustments are applied the other way around, that is first treating e_3 (which in this case entails adjusting *Other costs* and *Total costs*) and then e_1, we obtain the results in the column "e_3 *adjusted first*".

The differences in the results for the two different orders are quite large as are the adjustments themselves. If e_1 is treated first, this results in a moderate proportional downwards adjustment of *Profit* and *Total costs* to make them sum up to 950. When *Other costs* is adjusted next, the adjustment is very large because before adjustment *Other costs* was already larger than *Total costs – Wages* and since in the first step *Total costs* was reduced this discrepancy has become larger so that *Other costs* has to be reduced by more than 50%. For the other order in which e_3 is treated first, the adjustment to *Other costs* is only 10% but in this case we end up with a very large downwards adjustment of *Profit*.

Table 1. Example business record: prorating using two orders of application.

Variable	*Name*	*Unadjusted*	*e_1 adjusted first*	*e_3 adjusted first*
x_1	Profit	330	304	180
x_2	Employees	**25**	**25**	**25**
x_3	Turnover main	1000	922	922
x_4	Turnover other	30	28	28
x_5	Turnover	**950**	**950**	**950**
x_6	Wages	**550**	**550**	**550**
x_7	Other costs	200	96	220
x_8	Total costs	700	646	770

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

Banff Support Team (2008), Functional Description of the Banff System for Edit and Imputation. Technical Report, Statistics Canada.

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.

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 violations of balance edits by a uniform multiplicative adjustment to some variables involved in the edit.

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. The order in which the edit rules are treated can influence the result.

11. Variants of the method

1. Unweighted prorating
2. Weighted prorating

12. Input data

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

13. Logical preconditions

1. Missing values
    1. Edits with missing values cannot be handled by this method.
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. In the output data inconsistencies with respect to equality edits that existed in the input data are resolved.

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 – Minimum Adjustment Methods
3. Micro-Fusion – Generalised Ratio Adjustments

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

1.

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. Statistics Canada's generalised edit and imputation software Banff contains a routine PRORATE that provides an off-the-shelf, generalised prorating application. However, for specific applications the prorating calculations are not difficult to implement. So, without the availability of generalised prorating software, the application of prorating could be performed by an ad hoc implementation using general statistical packages with programming facilities such as R or SAS.

28. Process step performed by the method

GSBPM Sub-process 5.3: Review, validate and edit

Administrative section

29. Module code

Micro-Fusion-M-Prorating

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	10-12-2013	third version	Jeroen Pannekoek	CBS (Netherlands)
0.3.1	12-12-2013	preliminary release		
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 18:00