This module is part of the

Memobust Handbook

on Methodology of Modern Business Statistics

26 March 2014

Method: Stone's Method

Contents

General section

1. Summary

This Stone method is a method for data reconciliation. It adjusts data in order to satisfy a set of linear constraints. The adjustments made to the data are as small as necessary to remove all discrepancies. In adjusting the initial data the method of Stone uses information on the relative reliabilities of the initial data, in particular a covariance matrix. Data that are considered most reliable are modified least and vice versa. The Stone method yields a set of fully reconciled data, with minimum variance.

The method of Stone translates the reconciliation problem into a mathematical optimisation problem. From a mathematical perspective, the method of Stone is a weighted quadratic optimisation problem under linear conditions. The formulation of this problem is relatively easy to understand.

The solution of the model includes the reconciled data as well as its covariance matrix. Analytical expressions can be derived for both results.

2. General description of the method

2.1 Literature

A detailed description of the Stone method is given in the original paper, Stone (1942). In view of the extremely technical nature of this article, readers who are unfamiliar with the method are referred to the appendix in Wroe et al. (1999) and Eurostat (2008, section 14.4.2 “Lagrange method”). A mathematical derivation of the results is given in Sefton and Weale (1995).

The Stone method is widely researched in the literature. Several extensions are described. For instance for reconciliation problem with soft constraints, (hard and soft) nonlinear constraints (for instance ratios) and inequalities, see for instance Magnus et al. (2000).

2.2 Determining a covariance matrix

In practical applications of the method, a covariance matrix of the initial data is often unavailable. Therefore applications generally use estimates of relative variances. Relative variances have no intrinsic meaning, but the ratio of relative variances is an indicator of the reliability of figures relative to each other. There are several ad hoc methods for estimating relative variances. One method is to have a specialist estimate 95% confidence intervals and to use the interval sizes as an approximation for variances. Another method is to distinguish several categories, such as relatively unreliable, normally reliable and relatively reliable. All variables within the same group are assigned the same variance.

It is often desirable in practice for reconciliation to affect large values more than small values in an absolute sense. If this is what is intended, the following variances may be chosen:

$$Var(x_i) = \theta_i^{\,2} x_i^{\,2},$$

where θ_i is a parameter that depends on the reliability, or reliability category, of x_i.

Determining the correct ratios between the various variances is a process of trial and error in practice, which means that one particular ratio is chosen based on a degree of prior knowledge and simple assumptions (e.g., that variances are equal in the absence of prior knowledge), and then judging whether the results are acceptable. If not, the variances are modified.

In practice, in the absence of quantitative measures, all covariances are usually assumed to be zero, or, in other words, that the variables are assumed to be mutually independent.

3. Preparatory phase

4. Examples – not tool specific

This example is based on the greatly simplified supply and use tables, which belong to the national accounts, as shown in Table 1 and 2. The rows of Table 1 are related to the supply of products and services, and columns to the producing sectors. The first two rows of Table 2 show the demand for products and services, and the first two columns show the customer sectors. The grand total of the whole table is empty, since it was opted not to include it in the mathematical model. This grand total can be derived directly from the other totals.

There are only two sectors, industry and services, and two goods groups, industrial products and services. The economy depicted is moreover a closed one, since there is no trading with foreign countries. Other items ignored are taxes on products, subsidies, trade and transport margins, and all categories of final use other than consumption.

The constraints are that:

- total supply equals total use for industry and services (the column totals of Table 1 equal the first two column totals of Table 2);
- total supply equals total use for industrial products and services (row totals in Table 1 equal the first two row totals of Table 2).

In addition, the sums of the entries of Tables 1 and 2 must also equal its row and column totals.

Table 1. Supply

	Industry	Services	Total
Industrial products	700	300	1000
Services	100	400	500
Total	800	700	

Table 2. Use

	Industry	Services	Consumption	Total
Industrial products	50	190	860	1100
Services	170	100	180	450
Wages	450	350		800
Operating surplus	130	60		190
Total	800	700	1040	

Two constraints are not satisfied in the starting situations: total supply is unequal to total use for industrial products and services (the row totals of Table 1 are inconsistent with the first two row totals of Table 2). The variances are shown in Tables 3 and 4; they were chosen arbitrarily.

Table 3. Variances: supply table

	Industry	Services	Total
Industrial products	100	1000	1100
Services	1000	100	1100
Total	1100	1100	X

Table 4. Variances: use table

	Industry	Services	Consumption	Total
Industrial products	500	1000	1000	2500
Services	1000	1000	1000	3000
Wages	700	700		1400
Operating surplus	1200	1200		2400
Total	3400	3000	2000	X

Note that the row and column totals are not fixed, since their variance is greater than zero.

The figures are reconciled with the method of Stone. The reconciled values in Tables 5 and 6 satisfy all constraints. Small differences in the row sums in Table 6 are attributable only to rounding errors. The figures before reconciliation are shown in brackets.

Table 5. Table of reconciled supply values, rounded

	Industry		Services		Total	
Industrial products	705	(700)	318	(300)	1023	(1000)
Services	92	(100)	396	(400)	488	(500)
Total	797	(800)	714	(700)	1511	(1500)

Table 6. Table of reconciled use values, rounded

	Industry		Services		Consumption		Total	
Industrial products	33	(50)	164	(190)	827	(860)	1023	(1100)
Services	179	(170)	118	(100)	191	(180)	488	(450)
Wages	452	(450)	358	(350)			810	(800)
Operating surplus	133	(130)	74	(60)			207	(190)
Total	797	(800)	714	(700)	1017	(1040)	2527	(2540)

A covariance matrix is also derived for the reconciled figures. This covariance matrix is not diagonal, and there are also nonzero covariances. The variances are shown in Tables 7 and 8. The values are less than in the initial situation. The variances before reconciliation are shown in brackets.

Table 7. Variances for the table of reconciled supply values

	Industry	Services	Total
Industrial products	84 (100)	270 (1000)	280 (1100)
Services	277 (1000)	85 (100)	292 (1100)
Total	293 (1100)	289 (1100)	

Table 8. Variances for the table of reconciled use values

	Industry	Services	Consumption	Total
Industrial products	346 (500)	524 (1000)	463 (1000)	280 (2500)
Services	541 (1000)	523 (1000)	489 (1000)	292 (3000)
Wages	415 (700)	420 (700)		519 (1400)
Operating surplus	575 (1200)	591 (1200)		667 (2400)
Total	293 (3400)	289 (3000)	563 (2000)	

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

Eurostat (2008), *Manual of Supply, Use and Input-Output Tables*. Eurostat Methodologies and Working Papers.

Magnus, J. R., van Tongeren, J. W., and de Vos, A.F. (2000), National Accounts Estimation using Indicator Ratios. *The Review of Income and Wealth* **3**, 329–350.

Mantegazza S. and Di Leo, F. (2007), Integration of SUT/IOT into the National Accounts: The Italian experience. 16th International Conference on Input - Output Techniques, 2-6 July 2007, Istanbul, Turkey.

United Nations, Statistics Division (2000), *Handbook of National Accounting: Use of Macro Accounts in Policy Analysis*. Studies Methods, United Nations, New York.

Sefton, J. and Weale, M. R. (1995), *Reconciliation of national income and expenditure: balanced estimates for the United Kingdom, 1920-95*. Cambridge University Press, Cambridge.

Stone, J. R. N., Champernowne, D. A., and Maede, J. E. (1942), The Precision of National Income Accounting Estimates. *Reviews of Economic Studies* **9**, 111–125.

Wroe D., Kenny, P., Rizki, U., and Weerakoddy, I. (1999), *Reliability and Quality Indicators for National Accounts Aggregates*. Office for National Statistics (ONS), Document CPNB 265-1 for the 33rd meeting of the GNP Committee.
http://epp.eurostat.ec.europa.eu/pls/portal/docs/PAGE/PGP_DS_QUALITY/TAB47143266/RELIABILITY%20AND%20QUALITY%20INDICATORS%20FOR%20NATIONAL%20ACCOUNTS%20AGGREGATES.PDF

Specific section

8. Purpose of the method

The method is used for data reconciliation which is a specific process step used in the context of macro-integration (cf. “Macro-Integration – Main Module”).

9. Recommended use of the method

1. The method may be applied to any problem, in which consistency has to be achieved towards some set of equality constraints, which satisfies the preconditions in section 13.
2. Both positive and negative values are allowed. However, there is no way to constrain positive figures to remain positive and negative values to remain negative.
3. The method allows for exogenous variables, which are values that must remain unmodified.
4. The Stone method is relatively simple to program in R, Matlab and other packages. Quadratic programming (QP) solvers, such as CPLEX and XPRESS, can also be used.
5. The method may be applied to micro- or macrodata.
6. The method should be used to unbiased source figures. All source figures are consistent with their definitions. They are therefore already adjusted for systematic errors (nonresponse errors, measurement errors, processing errors and conceptual differences). Any errors in the input data (as mentioned in section 12) will propagate to the results.

10. Possible disadvantages of the method

1. Zero values remain zero
2. The variances may be chosen arbitrarily

11. Variants of the method

1.

12. Input data

1. Ds-input1 = a data set (microdata or macrodata) (required)

13. Logical preconditions

1. Missing values
    1. In Ds-input1 individual missing data values are not allowed.
2. Erroneous values
    1.
3. Other quality related preconditions
    1.
4. Other types of preconditions

1. The constraints (section 14.2) must not be mutually conflicting.
2. Only equality constraints. Inequality constraints, such as ‘revenue > 100 × number of active employees’ are therefore not supported. Since non-negativity is a special case of inequality constraint, it cannot be modelled.
3. Only linear constraints.

14. Tuning parameters

1. A covariance matrix (Required). The relative variance determine which of the variables are adjusted the most.

Remark 1: When all variables are equally reliable an identity matrix may be used.

Remark 2: This covariance is usually called the ex-ante covariance matrix. It differs from the ex-post covariance matrix, as mentioned in section 21.1.

2. Constraints (Required). These specify the linear constraints that should be satisfied.

15. Recommended use of the individual variants of the method

1.

16. Output data

1. Ds-output1 = a dataset with reconciled (micro- or macrodata) sub-annual time series.

17. Properties of the output data

1. The output data (Ds-output1) satisfy all linear constraints (section 14.2).
2. The ex-post variances (in section 21.1) can be as most as large as the corresponding ex-ante variances (in section 14.1).
3. The amount of adjustment is at least as possible.

18. Unit of input data suitable for the method

Processing full data sets

19. User interaction - not tool specific

1. Before execution of the method, the tuning parameters and input datasets must be specified.
2. During operation no user interaction is needed, but inspection of quality indicators and subsequent adjustment of tuning parameters and recurrent use is optional.
3. After use of the method the quality indicators and logging should be inspected.

20. Logging indicators

1. The run time of the application.
2. Characteristics of the input data, for instance problem size, the largest discrepancies of the input data towards the constraints.

21. Quality indicators of the output data

A quality indicator of the output (Ds-output1) is:

1. A covariance matrix, corresponding to Ds-output1 (usually called the ex-post covariance matrix).

Quality indicators of the method are:

2. The most important quality indicator is *how* the figures (in Ds-input1) were adjusted. Attention may be focused on relative or absolute differences. Because of the relationships between the various variables in the system, the differences must be examined in their mutual context. A quantitative measure for the degree of inconsistency in the data before reconciliation is the value of a weighed sum of the squared reconciliation adjustments. A high value implies many large adjustment were needed to achieve consistency.
3. Another quality aspect is accuracy. The method of Stone gives reconciled figures with minimum variance, assuming the variance of the figures to be reconciled are given. The diagonal entries of the ex-post covariance matrix (section 21.1) give information about the relative reliability of the reconciled results. Comparison with the ex-ante covariance matrix (section 14.1) yields information about how the reconciliation reduces the data variance. The non-diagonal entries of the ex post covariance matrix (section 21.1) yield information about inter-variable correlations introduced by reconciliation.

Remark 1: This process can become extremely complicated with very large numbers of variables or internal relationships, in which case it may be simpler to analyse the differences before reconciliation (i.e., the constraint violations), as opposed to the reconciliation adjustments.

Remark 2: A need for many large reconciliation adjustments may indicate biased source data, meaning that the model conditions were not satisfied, and therefore that the method should not have been applied.

Remark 3: An important quality indicator of the method *implementation* is *whether* the figures are successfully reconciled. To this end, the remaining differences can be calculated on all linear constraints (section 14.2). Numerical error will generally cause these differences to deviate slightly from zero, which is not usually a problem as long as the differences are less than a certain threshold value.

22. Actual use of the method

1. The method of Stone adapted by National Statistical Offices in the compilation of National Accounts, for instance by ISTAT (Mantegazza and Di Leo, 2007).

Interconnections with other modules

23. Themes that refer explicitly to this module

1. Macro-Integration – Main Module

24. Related methods described in other modules

1. Macro-Integration – RAS

2. Macro-Integration – Denton’s Method
3. Macro-Integration – Chow-Lin Method for Temporal Disaggregation

Remark 1: The Stone method is more general than the RAS method. The RAS method adjusts the entries of an matrix to achieve consistency with given row and column totals. The method of Stone however does not need the precondition that the data can be represented in a two-dimensional matrix. Furthermore, the method of Stone allows for different types of constraints than the alignment to row and column totals. And a third difference is that the method of Stone allows for differences of reliability of the source data, while the RAS method does not. However, from a technical point of view, the RAS method is easier to apply than the Stone method. The RAS method is an easy iterative scaling procedure, while the Stone method requires the computation of the solution of a least square problem.

Remark 2: In comparison with the Denton method, the Stone method is less specific. The Denton method is meant for achieving consistency between data of different frequencies (for instance quarterly data that has to comply with annual data), while the Stone method does not include a time component.

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

1. Generalised Regression
2. Quadratic programming under linear constraints

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

1. GSBPM phase 6.2 “Validate Outputs”

27. Tools that implement the method described in this module

1.

28. Process step performed by the method

Data reconciliation

Administrative section

29. Module code

Macro-Integration-M-Stone

30. Version history

Version	Date	Description of changes	Author	Institute
0.1	31-03-2011	first version	Jacco Daalmans	CBS
0.2	27-01-2012	second version	Jacco Daalmans	CBS
0.3	21-06-2013	third version	Jacco Daalmans	CBS
0.3.1	06-09-2013	preliminary release		
0.3.2	09-09-2013	page numbering adjusted		
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	26-3-2014 13:25