Macro Integration

Data reconciliation

10

Reinier Bikker, Jacco Daalmans and Nino Mushkudiani

Statistical Methods (201104)

Statistics Netherlands

The Hague/Heerlen

Explanation of symbols

.	= data not available
*	= provisional figure
**	= revised provisional figure
x	= publication prohibited (confidential figure)
–	= nil or less than half of unit concerned
–	= (between two figures) inclusive
0 (0,0)	= less than half of unit concerned
blank	= not applicable
2010–2011	= 2010 to 2011 inclusive
2010/2011	= average of 2010 up to and including 2011
2010/'11	= crop year, financial year, school year etc. beginning in 2010 and ending in 2011
2008/'09–2010/'11	= crop year, financial year, etc. 2008/'09 to 2010/'11 inclusive

Due to rounding, some totals may not correspond with the sum of the separate figures.

Publisher
Statistics Netherlands
Henri Faasdreef 312
2492 JP The Hague

Prepress
Statistics Netherlands - Grafimedia

Cover
TelDesign, Rotterdam

Information
Telephone +31 88 570 70 70
Telefax +31 70 337 59 94
Via contact form: www.cbs.nl/information

Where to order
E-mail: verkoop@cbs.nl
Telefax +31 45 570 62 68

Internet
www.cbs.nl

© Statistics Netherlands, The Hague/Heerlen, 2011.
Reproduction is permitted. 'Statistics Netherlands' must be quoted as source.

60165201104 X-37

Table of contents

1. Introduction to the subtheme

1.1 General description and reading guide

1.1.1 Description of the subtheme

Macro integration has two objectives. The first is to facilitate analysis of the interrelationships in data by organizing economic information into an accounting framework. The second is to make more accurate estimates of economic reality through the confrontation or reconciliation of the various statistical information contained in a framework of this kind. An economic accounting framework is defined by a set of variables and a set of relationships between them.

Macro integration first 'translates' the source data to comply with the correct definitions and then identifies and adjusts for major measurement errors. This process entails confronting the various variables through their relationships. The data generally come from a wide variety of sources. Some variables may be obtained from external sources, or through sampling, but, where no suitable source exists, variables may be based on model estimates, or 'expert guesses'.

Both bias and sampling errors generally have to be detected and reduced during the data reconciliation process. The formal data reconciliation methods described below are concerned mainly with error, and are unsuitable for detecting and eliminating bias. It is hard to differentiate between bias and sampling errors in practice. It is usually necessary to resort to elimination by hand of the largest differences, and to equally distribute the mass of smaller discrepancies through modelling.

1.1.2 Problems and solutions

The literature refers to various formal data reconciliation methods, each with its own origins. There is a correspondingly great variety in applicability, interpretability and generality. All the methods covered below are suitable for computer implementation.

The simplest methods were devised at a time before powerful computers were widely available. An example is the *RAS method*, which allows the entries of a rectangular matrix to be aligned with a set of row and column totals. This method is described in Chapter 2.

There are more general methods, with a better statistical foundation, that estimate reconciled results from source data while complying with certain constraints, in accordance with a specific procedure. Almost all the common methods can be classified as *generalized least-squares methods*, which all have quadratic error terms in the objective function. The best statistical estimate corresponds with the optimum value of the objective function. Different additional assumptions give rise to specific model variants.

The assumption that the variables are mutually independent and that there are linear constraints that should hold exactly (without an error term) leads to *Stone's method*, which is one of the older (1942) and most rudimentary of the least-squares methods. This method is described in Chapter 3. There are various possible extensions to Stone's method and two important ones that are described in Chapter 4 incorporate 'soft' constraints and ratio constraints, respectively.

Data reconciliation can also be extended with a temporal component, which involves using data from multiple time periods simultaneously. Data with unequal frequencies, such as quarterly and annual, are allowed. Relationships may occur within a single specific time interval, but also between different periods, an example of which is that the sum of four quarters equals the value for the year. The same kinds of inconsistencies can occur between variables observed on, say, a quarterly and an annual basis as between different variables of the same period.

Two terms are mentioned frequently in the literature. The process to achieve consistency between data that are published at a high frequency and data that are published at a low frequency is usually referred to as *benchmarking*. Closely related to benchmarking is *temporal disaggregation,* which are methods for deriving high frequency data from low frequency data by means of indicators.

The Chow-Lin (1971) method can be used in both application areas. It applies a least square estimator to derive a high frequency (e.g. quarterly) series from one or more related series (indicators) in a way that makes the high frequency series correspond with a low frequency (e.g. annual) series. This method finds the solution that best fits the set of indicators. However, the method is less suited to fast-changing regression relationships in time, for which it creates step problems, with disproportionately large transitions from the fourth quarter of one year to the first quarter of the next. The authors are unaware of any application of the Chow-Lin method within Statistics Netherlands, and it is therefore not covered in detail in this method's review.

Chapter 5 presents the univariate *Denton method*. This method adjusts one high frequency series to fit one low frequency series, without creating step problems. The low frequency series are exogenous, which means that they are not modified.

The Denton method preserves as much as possible the trend exhibited in the high frequency series, and not the levels. The motivation for this principle is that first differences (the difference between two successive periods) of the high frequency indicator are the only source of information about the short term movements in the series to be estimated, while the levels and the long term trend of the low frequency series are assumed to be more reliable than those of the high frequency indicator.

The Denton method accordingly combines the information contained in the first differences of the high frequency data series with information about the level and the long term trend from the low frequency series. From this combination it then estimates a new high frequency series that corresponds with the low frequency series.

A variant of this method known as the Cholette matrix method is used in Statistics Netherlands for benchmarking the main economic series after seasonal adjustment.

Chapter 6 discusses a rudimentary form of the *multivariate Denton method*. This method is able to align a series of high frequency *data sets* with a series of low frequency data sets. In this case linear constraints apply within one period and between the various frequencies. The multivariate Denton method can also be extended to handle soft constraints and ratio constraints. A description of this extended model is given in an other contribution to the Methods Series.

1.2 Scope and relationship with other themes

We selected the most general and rudimentary version of each of the available methods. We also describe some possible general extensions. References are made where possible to the literature rather than repeating the subject matter here.

There are no direct relationships with other themes in this Methods Series.

1.3 Place in the statistical process

The data reconciliation methods described below are primarily designed for integrating observed economic quantities into accounting frameworks. The data reconciliation process invariably occurs at the end of the economic statistics production chain.

However, problems that are amenable to the same or similar techniques may also occur elsewhere.

2. RAS / IPF

2.1 Short description

The starting point of the RAS method is a matrix whose entries are inconsistent with the row and column totals. The result is a new set of entries that are consistent with the row and column totals. The totals remain the same, while the entries change. As much as possible of the structure of the entries is retained, in the sense that relatively small values are kept relatively small, and vice versa.

The algorithm is a fairly simple iterative procedure, in which the rows and columns are corrected in turn. There is a more specific description in section 2.3. The RAS method is also known as iterative proportional reconciliation (IPF), or biproportional reconciliation in the two-dimensional case.

2.2 Applicability

The RAS method is suited to problems in which the aim is to make figures in a rectangular matrix consistent with row and column totals. This method is only interesting if it is possible to represent the data in one rectangular matrix, in which the entries of the matrix are subject to change, while its row and column totals are fixed to their initial values. The table does not have to be square: the number of rows may differ from the number of columns.

There is no way of differentiating between the hardness of different variables in the entries of the matrix. It is therefore impossible to make sure that a given value in the entries of the matrix is to undergo minimal change. The adjustment of the entries of the matrix always happens to be (multi)proportional to the row and column totals, in order to preserve as much as possible of the table structure.

There is likewise no facility to impose constraints other than the row and column totals. It is therefore impossible to prescribe that the sum of two variables in the matrix equals a third variable in the matrix, even if the condition was met in the initial situation.

The RAS method is very suitable for updating the Input Output Tables of National Accounts, for instance. At the current time, *t*, the row and column totals are fixed, since they have to be consistent with another statistic: the supply and use table. Figures about the entries of the matrix are available only up to and including $t - 1$. Updating involves modifying the entries of the matrix at $t - 1$ in such a way as to make them consistent with the row and column totals at time *t*, also preserving the structure at $t - 1$ as much as possible.

A technical description of the conditions of the method is given in Bacharach (1970). An overall description of the conditions, which avoids technical details, is given below.

1) Initial estimates are available for each variable (in the entries of the matrix and the row and column totals)[1];

2) The source figures can be represented in one table.

3) All values are non-negative.

4) Rows and columns consisting entirely of zeros do not occur in combination with a nonzero row or column total.

5) The sums of the row and column totals are equal.

2.3 Detailed description

This section describes the algorithm. See Bacharach (1970) for an explanation. The essence is that rows and columns are rescaled in turn to equal the row and column totals. The starting point is an $n \times m$ matrix A, and required row and column totals: r_i and k_j.

Step 0 Initialize: $t = 0$ and $A^{(t)} = A$;

Step 1 For each row i calculate the correction factor

$$c_i^{(t)} = \frac{r_i}{\sum_{j=1}^{m} A_{ij}^{(t)}}. \tag{2.1}$$

The correction factor is the ratio between the required and the current row totals. Multiply each matrix entry in row i by the correction factor. This produces a matrix $A^{(t+1)}$ for which the sums of the rows equal the required row totals. The method stops if all the column totals also equal the required column totals (or if the discrepancy is acceptably small). Otherwise continue with Step 2 and take $t = t + 1$.

Step 2 For each column j calculate the correction factor

$$c_j^{(t)} = \frac{k_j}{\sum_{i=1}^{n} A_{ij}^{(t)}}. \tag{2.2}$$

The correction factor is the ratio between the required and the current column totals. Multiply each entry in column j by the correction factor. This produces a matrix $A^{(t+1)}$, of which the sums of the columns equal the required column totals. Stop, if all row totals also equal the required totals (or if the discrepancy is acceptably small). Otherwise return to Step 1 and take $t = t + 1$.

Step 1 of the algorithm makes the rows consistent and Step 2 does the same for the columns. The sequence is arbitrary, and Step 2 could just as well precede Step 1.

This algorithm always converges to a solution, provided all conditions given in Bacharach (1970) are satisfied. These conditions closely resemble conditions 1) to

[1] If no information is available for the inner part of the matrix, condition 1) can still be satisfied by filling it with ones. The reconciled figures are then biproportional to the marginal totals.

5), inclusive, above, but are not identical. In order to avoid technical details, the exact conditions are not described in this document.

2.4 Example

The entries of the matrix in Table 2.1 below are inconsistent with the associated row and column totals. The RAS method is applied in order to produce a consistent table.

Table 2.1. Initial situation

2	4	12
2	4	6
9	9	**18**

The rows are modified first. The sum of the entries in the first row is six, while the row total is twice as large. The entries of the matrix must therefore be multiplied by two. The second row of the matrix entries is consistent with the row total and is therefore left unmodified. The result of the above is Table 2.2. The matrix entries of this table are consistent with the row totals, but not with the column totals.

Table 2.2. First intermediate result

4	8	12
2	4	6
9	9	**18**

The next step of the algorithm modifies the columns. The sum of the first column is six and the column total is nine. The entries in the first column must therefore be increased by a factor of 9/6, or 1.5. The sum of the entries in the second column is 12, while the column total is 9. The two entries of the matrix must therefore be multiplied by 9/12, or 0.75, which produces Table 2.3. This table is entirely consistent. The algorithm stops. (The algorithm would continue with rows if the table were still inconsistent.)

Table 2.3 Final result

6	6	12
3	3	6
9	9	**18**

2.5 Characteristics

All the reconciled values are non-negative. A zero entry of the matrix is not modified. Conversely, it is possible for a positive entry of the matrix to be modified to zero (but only because of a zero row or column total).

The result of the RAS method applied to matrix A is the same as the result of the following optimization problem:

$$\text{Min}_{A^*} \sum_{i=1}^{n} \sum_{j=1}^{m} A_{ij}^{*} (\ln A_{ij}^{*} - \ln A_{ij}), \tag{2.3}$$

$$\text{given} \sum_{i=1}^{n} A_{ij}^{*} = k_j, \; j = 1,\ldots,m \tag{2.4}$$

$$\sum_{j=1}^{m} A_{ij}^{*} = r_i, \; i = 1,\ldots,n \tag{2.5}$$

where n is the number of rows, m is the number of columns, k_j is a column total and r_i is a row total.

The multipliers c_i and c_j in (2.1) and (2.2) have an economic interpretation that depends on the practical application involved, such as a substitution effect and an effect that reflects the increased production of an industry as a whole. There is additional information in Bacharach (1970).

2.6 Quality indicators

The most important quality indicator is *how* the figures are adjusted. Relative or absolute differences may be explored. Because of the relationships between the various entries of matrix A, the differences must be examined in their mutual context. It would be possible, for example, to explore how the ratios between matrix cells and the row and column totals change in the reconciliation process.

The RAS method attempts to preserve these ratios as much as possible. If a ratio has to change in the reconciliation process nonetheless, it is advisable to review the suitability of RAS.

Special attention is needed for zeros, both before and after reconciliation. The RAS method may create zeros if a row or column total is zero, but cannot adjust existing zeros entries of the matrix. In either case it must be verified that the data set has the correct structure.

3. Stone's method

3.1 Short description

Stone's method modifies a set of figures as little as possible while satisfying certain constraints. Unlike RAS, the figures do not have to be presented in a rectangular table.

The results of the method may be driven by reliability weights (for instance variances). The weights can be devised to ensure that the figures deemed most reliable are modified least. The method provides reliability weights of the consistent values as output.

From a mathematical perspective, this is a weighted quadratic optimization problem under linear conditions. Analytic expressions exist for solving this problem. The Lagrange multiplier method can be used to derive the solution.

3.2 Applicability

The method is interesting in situations where a set of figures must be modified as little as possible while satisfying certain constraints. An extremely simple example of a constraint is the requirement for one variable to equal another.

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.

The results of the reconciliation method may be, but are not necessarily, driven by the use of reliability weights. Furthermore, the method allows exogenous variables, which are values that must remain unmodified.

The conditions are as follows.

1) Availability of figures. Source figures and an associated relative variance or covariance matrix are available.
2) Unbiased source figures, which means that all source figures are consistent with the definitions. They are therefore already adjusted for errors (nonresponse errors, measurement errors, processing errors and conceptual differences), since the mathematical method is not intended for adjusting for errors. Any errors in the input data will propagate to the results.
3) 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. It hardly needs to be added that the constraints must not be mutually conflicting.
4) Only linear constraints. Therefore, nonlinear constraints cannot be modelled. Ratios, such as 'value / volume = price', are nonlinear constraints. Chapter 4 presents an extension of this method that is able to handle a certain type (ratios) of nonlinear constraint.

3.3 Detailed description

3.3.1 Method

In view of the extremely technical nature of Stone's (1942) article, readers who are unfamiliar with the method are referred to the appendix in Wroe et al. (1999). A mathematical derivation of the results is given in Sefton and Weale (1995).

As stated in section 3.1, problem described by Stone is a quadratic optimization problem. The mathematical formulation is

$$\text{Min}_{x^*}\ (x^* - x)' V^{-1} (x^* - x), \tag{3.1}$$

$$\text{Given}\ \ Ax^* = b. \tag{3.2}$$

Where x is an n-dimensional column vector comprising the data. The particular arrangement of the data within this vector is irrelevant. The solution x^* is also an n-dimensional column vector. The objective function (3.1) minimizes the weighted quadratic difference between x^* and x. Weighting is carried out using with a matrix V^{-1}, which is the inverse of the relative covariance matrix. Additional information about this matrix is given in section 3.3.2. The boundary condition (3.2) forces x^* to comply with a system of linear constraints. A is the coefficient matrix of this system and the vector b comprises the coefficients on the right-hand side of the constraints.

In the special case of a diagonal covariance matrix (all x_i independent), (3.1) is equivalent to:

$$\text{Min}_{x^*} \sum_{i=1}^{n} \frac{1}{v_{ii}} (x_i^* - x_i)^2, \tag{3.3}$$

where v_{ii} represents the variance of x_i.

The expressions of the solutions are

$$x^* = x + VA'(AVA')^{-1}(b - Ax), \tag{3.4}$$

$$V^* = V - VA'(AVA')^{-1}AV. \tag{3.5}$$

V^* is the covariance matrix of x^*. A property of the diagonal entries of this matrix is that they are not greater than the corresponding entries in V. In other words, reconciliation increases reliability. The inverse in (3.4) can be determined only if the constraints are mutually independent. A coefficient matrix of a system of independent constraints can be obtained by performing Gaussian elimination on A and eliminating zero rows. An alternative is to use the Moore-Penrose inverse: see Knottnerus (2003, p.331).

Stone's method is frequently applied in the national accounts for extremely large matrices. Inverting a matrix in formula (3.4) can then present a computational problem. It is advisable to use sparse matrices.

Knottnerus (2003) has described an iterative Kalman method that in theory is suitable for inverting extremely large matrices.

3.3.2 The covariance matrix

Matrix V in (3.4) comprises relative (co)variances. The proportions of the numbers in this matrix determine the values that will be modified the most. A question that arises is how to determine this matrix. Ideally, the matrix would contain true (co)variances, but in practice, however, these are often unavailable. Practical 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 the variables in a given category acquire the same weights (θ in (3.6) below).

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:

$$Var(x_i) = \theta_i^{\,2} x_i^{\,2}, \tag{3.6}$$

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 mutually independent.

3.3.3 Software

Winadjust is available from Statistics Netherlands for the application of Stone's method. See Van Dalen (2002) for a detailed manual.

Winadjust is suitable only for the specific reconciliation problem in which the entries of a table are to be made consistent with the row and column totals. Under normal circumstances the row and column totals are fixed and only the entries are modified. A workaround, if either the row or column totals are not fixed, is to move the margins into the inner part of the matrix. Readers are referred to Van Dalen (2002) for a description of this technique.

In Winadjust the weights can be read in from a file or created automatically.

The program offers the option that the programme stops if a value changes sign (from positive to negative or vice versa).

Besides reconciliation, Winadjust can also be used to round results consistently to whole numbers, in the sense that the entries of the matrix and the row and column totals remain consistent after rounding.

Stone's method is also relatively simple to program in Matlab and other packages. Quadratic programming (QP) solvers, such as CPLEX and XPRESS, can also be used.

3.4 Example

This example is based on the greatly simplified supply and use tables shown in Tables 3.1 and 3.2. The rows of Table 3.1 are related to the supply of products and services, and columns to the producing sectors. The first two rows of Table 3.2 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 3.1 equal the first two column totals of Table 3.2);
- total supply equals total use for industrial products and services (row totals in Table 3.1 equal the first two row totals of Table 3.2).

Needless to say, the sums of the entries of Tables 3.1 and 3.2 must also equal the row and column totals.

Table 3.1: Supply

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

Table 3.2: 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 3.1 are inconsistent with the first two row totals of Table 3.2). The variances are shown in Tables 3.3 and 3.4; they were chosen arbitrarily.

Table 3.3: Variances: supply table

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

Table 3.4: 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 Stone's method. The reconciled values in Tables 3.5 and 3.6 satisfy all the constraints. Small differences in the row sums in Table 3.6 are attributable only to rounding errors. The figures before reconciliation are shown in brackets.

Table 3.5: 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 3.6: 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 3.7. and 3.8. The values are less than in the initial situation. The variances before reconciliation are shown in brackets.

Table 3.7: 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 3.8: 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)		

3.5 Quality indicators

The most important quality indicator is *how* the figures were adjusted. Attention may be focused on relative or absolute differences, depending on the properties of the weights entered in the covariance matrix *V*. Because of the relationships between the various variables in the system, the differences must be examined in their mutual context. 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, as opposed to the adjustments.

A quantitative measure for the degree of inconsistency in the data before reconciliation is the value of the objective function (3.1). The sum of all reconciliation adjustments increases in size in line with the value of the objective function. A need for many 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.

Another quality aspect is accuracy. Stone's method gives reconciled figures with minimum variance, assuming the variance of the figures to be reconciled are given. The ex post covariance matrix of the reconciled values can be calculated with (3.5). The diagonal entries of this matrix have information about the relative reliability of the reconciled results. Comparison with the covariance matrix used in the objective function yields information about how the reconciliation reduces the data variance. The nondiagonal entries of the ex post covariance matrix yield information about intervariable correlations introduced by reconciliation.

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. 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.

4. Extensions to Stone's method

4.1 Short description

The literature describes several extensions to Stone's method, see e.g. Magnus and others (2000) and Boonstra (2006). This chapter considers two of these extensions.

1) *Reconciliation problems with soft constraints*

   A constraint is deemed soft if it is acceptable for it to be satisfied 'approximately' (whereas hard constraints have to be satisfied exactly). An example is that a (forecast) quarterly value of some variable must be approximately equal to the previous year's value.

   It is possible to incorporate soft constraints into the mathematical reconciliation model, while taking account of differences in reliability.

2) *Reconciliation problems with constraints in the form of a ratio*

   An assumption about a relationship between two variables can be modelled mathematically as a ratio. An example of a ratio constraint is that the ratio of milk exports to production must be approximately 0.2. Ratio constraints can be either hard or soft.

   It is possible to incorporate ratio constraints in the mathematical reconciliation model, again while taking account of differences in reliability.

4.2 Applicability

This method is interesting for reconciliation problems in which substantial professional knowledge is available, in particular if this knowledge is about relationships between variables. Examples include supply-use ratios, labour income shares, growth rates (t divided by $t - 1$) and prices (a price is a change in value divided by a change in volume). This knowledge can be converted into binding and nonbinding ratio or other constraints.

With one exception, the conditions given in section 3.2 are applicable here too. The one exception has to do with linear constraints. This chapter describes a reconciliation method that is also suitable for nonlinear (ratio) constraints.

4.3 Detailed description

Magnus and others (2000) describe a reconciliation method for a model that is able to incorporate soft constraints and ratios. The extremely technical nature of their description makes it difficult to comprehend. A mathematical explanation is given in Boonstra (2006) and a description 'in words' is given in United Nations (2000).

Descriptions are given below of methods for achieving the above two methodological extensions.

4.3.1 Soft constraints

A mathematical formulation of an optimization problem with hard and soft conditions is:

$$\text{Min}_{x^*}\ (x^* - x)' V^{-1}(x^* - x) + (A_1 x^* - b_1)' \Sigma^{-1}(A_1 x^* - b_1), \tag{4.1}$$

$$\text{Given } A_2 x^* = b_2. \tag{4.2}$$

In which (4.2) is a system of hard constraints. The soft constraints are shown in the system

$$A_1 x^* \text{ where} \tag{4.3}$$

$$\text{E}(A_1 x^*) = b_1 \text{ and} \tag{4.4}$$

$$\text{Cov}(A_1 x^*) = \Sigma. \tag{4.5}$$

The expressions for the solution of (4.1) and (4.2) are

$$x^* = x + VA'(AVA' + \bar{\Sigma})^{-1}(b - Ax) \tag{4.6}$$

and

$$V^* = V - VA'(AVA' + \bar{\Sigma})^{-1} AV, \tag{4.7}$$

where

$$A = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \text{ and } b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}. \tag{4.8}$$

(4.6) and (4.7) show a $\bar{\Sigma}$ square matrix with a dimension equal to the number of constraints. This matrix has the form

$$\bar{\Sigma} = \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix}, \tag{4.9}$$

the upper rows correspond with the soft constraints, and contain the associated covariance matrix Σ. The lower rows are associated with the hard constraints, and contain only zeros, since all (co)variances of the hard constraints are zero.

4.3.2 Ratios

Ratios are nonlinear, whereas all hard and soft conditions (4.2) and (4.3) are linear. However, it is possible to add conditions in the form of ratios to the model if they are first linearized. Magnus and others (2000) and Boonstra (2006) show that the ratio x_1 / x_2, with

$$\text{E}\left(x_1 / x_2\right) = r \text{ and} \tag{4.10}$$

$$\text{Var}\left(x_1 / x_2\right) = \sigma_R^2, \tag{4.11}$$

is equivalent to the soft constraint:

$x_1 - r\,x_2$, with (4.12)

$E(x_1 - r\,x_2) = 0$ and (4.13)

$$\mathrm{Var}(x_1 - r\,x_2) = \sigma_R^2\left(\sigma_{x_2}^2 + \mathrm{E}(x_2)^2\right). \quad (4.14)$$

Section 4.3.1 showed that soft constraints can be incorporated into the model. An assumption made in the above derivation is that R and x_2 are independent.

The same problem is also dealt with in Knottnerus (2003). Although different mathematical methods are used, the expressions for the solutions are the same.

4.4 Example

This example continues the example in section 3.4. The conditions given in section 3.4 are augmented with a modelled ratio.

The ratio is

$$R = \frac{\text{Use of industrial products by industry}}{\text{Total use by industry}}. \quad (4.15)$$

It is assumed that the predicted value of this ratio is 0.063: which is approximately the same value as before the reconciliation (50/800), see Table 3.2. The variance of R is assumed to be 0.0001. This value is chosen based on a 95% margin of uncertainty, amounting to 0.02. Table 3.6 shows that if the ratio is not included in the model, the value of the ratio after reconciliation will be 0.041 (= 33/797). Tables 4.1 and 4.2 show how the results change if the ratio is included in the model. Reconciled values calculated without the ratio constraint are shown in brackets. These results are from Tables 3.5 and 3.6.

Table 4.1: Supply – reconciled.

	Industry		Services		Total	
Industrial products	705	(705)	320	(318)	1025	(1023)
Services	93	(92)	396	(396)	489	(488)
Total	798	(797)	716	(714)	1514	(1511)

Table 4.2: Use – Reconciled

	Industry		Services		Consumption		Total	
Industrial products	48	(33)	158	(164)	820	(827)	1025	(1023)
Services	174	(179)	121	(118)	193	(191)	489	(488)
Wages	449	(452)	360	(358)			809	(810)
Operating surplus	128	(133)	77	(74)			205	(207)
Total	798	(797)	716	(714)	1014	(1017)	2528	(2527)

The addition of one ratio causes almost all the values to change, albeit mostly only slightly. The value of the use of industrial products by industry changes most, which is logical, in that the variable is in the numerator of the ratio. The value of the ratio after reconciliation is: 0.060 (= 48/798). This differs only slightly from 0.063, the predicted value of the ratio. Including the ratio in the reconciliation model therefore has a conspicuous influence on the results.

4.5 Characteristics

It is possible from a methodological perspective to model ratios and soft constraints. This facility allows professional knowledge, which is not based on observation, to be incorporated in the model.

From a technical perspective it is disputable whether professional knowledge should be incorporated in a reconciliation model, but an argument in favour of doing so is that the information can be deemed reliable and therefore adds value. Results that are contrary to professional insight are undesirable.

Conversely, a reason for leaving professional knowledge outside the model is that it impedes checking the observed figures for errors. If the reconciled figures clearly run counter to professional knowledge, the most likely reason is the presence of errors in the observed figures. These errors may remain undetected if professional knowledge plays a part in the reconciliation.

4.6 Quality indicators

The most important quality indicator is *how* the figures were adjusted. Attention may be focused on relative or absolute differences, depending on the properties of the weights entered in the covariance matrix V. Because of the relationships between the various variables in the system, the differences must be examined in their mutual context. 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, as opposed to the adjustments.

The extensions to Stone's method described in this chapter can be incorporated simply in the checking of the changes. In this case the ex-post values of the ratios, and the ex-post differences on soft relationships are involved.

A quantitative measure for the degree of inconsistency in the data before reconciliation is the value of the objective function (4.1). The sum of all reconciliation adjustments increases in size in line with the value of the objective function. A need for many 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.

Another quality aspect is accuracy. Stone's method gives reconciled figures with minimum variance, given the variance of the figures to be reconciled. The ex-post covariance matrix of the reconciled values can be calculated with (4.7). The diagonal entries of this matrix have information about the relative reliability of the reconciled results. Comparison with the covariance matrix used in the objective

function yields information about how the reconciliation reduces the data variance. The nondiagonal entries of the ex-post covariance matrix yield information about intervariable correlations introduced by reconciliation.

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. 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.

5. Univariate Denton method

5.1 Short description

The problem that the Denton method (Denton 1971) addresses occurs frequently in analysing economic time series. The problem is of aligning high frequency (e.g. quarterly) time series with low frequency (e.g. annual) series from a different source. This means, for example, that the sum of the figures of four quarters must equal the annual figure. It is assumed below, without loss of generality, that a quarterly series is being aligned with an annual series. The annual series is furthermore exogenous: it will not be modified, whereas the quarterly series will be modified.

The Denton method attempts to preserve the first-order differences in the quarterly series, as opposed to the levels. This assumption is also referred to as the 'movement preservation principle'. The underlying assumption is that the changes in the quarterly series are measured more accurately than the levels.

The optimal solution is the one with the least average modification to the first-order differences over the entire period of the series. This means that the value of a quarterly figure is determined not only by the corresponding annual figure but also by the annual figures before and after the corresponding year. In this way the method avoids creating a large step change between the last quarter of one year and the first quarter of the next.

5.2 Applicability

The aim of the method is to align a quarterly series with an annual series while preserving as much as possible the original quarter-to-quarter changes (the movement preservation principle).

An example of an application is the alignment of a seasonally sensitive quarterly series with an annual series, where the seasonal patterns have to be preserved as much as possible. The method is particularly interesting if it is important to avoid large step changes between the last quarter of one year and the first quarter of the next. It is of little importance whether the differences between quarterly and annual figures arise from sample error or statistical manipulations, such as seasonal adjustment.

The method is subject to the following conditions.

1. Source figures (annual and quarterly) are available. This implies that quarter-to-quarter changes are also known.

2) Unbiased source figures. This means, for example, that the definitions of the source figures for years and quarters correspond with those of the national accounts, and are therefore already adjusted for sampling errors, nonresponse errors, measurement errors, processing errors and conceptual differences.

5.3 Detailed description

We describe in this section the Denton method when quarterly data are from one source and annual data from a different source (Denton 1971). The Denton method is derived from Stone's method below.

The quarterly data is represented by a column vector x and the annual data by a column vector b,

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_m \end{pmatrix}, \tag{5.1}$$

where $n = 4m$. We want the following to apply:

$$\sum_{j=4(k-1)+1}^{4k} x_j = b_k, \quad k = 1, \ldots, m. \tag{5.2}$$

These equalities will almost never be satisfied in the initial situation. We therefore seek a vector x^* that in accordance with some criterion is close to the original x and satisfies (5.2). Stone's method minimizes the quadratic differences between the original vector x and the modified vector x^*. As in section 3 we obtain a convex optimization problem:

$$\min_{x^*} (x^* - x)' V^{-1} (x^* - x), \tag{5.3}$$

$$\text{given} \sum_{j=4(k-1)+1}^{4k} x_j^* = b_k, \quad k = 1, \ldots, m \tag{5.4}$$

Here V is a symmetrical, nonsingular matrix. This quadratic problem, see also (3.3), can be solved by means of Lagrange multipliers:

$$(x^* - x)' V^{-1} (x^* - x) - \lambda' (b - Ax^*), \tag{5.5}$$

with

$$\lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \cdot \\ \cdot \\ \cdot \\ \lambda_m \end{pmatrix} \text{ and } A' = \begin{pmatrix} j & 0 & \times\times\times & 0 \\ 0 & j & \times\times\times & 0 \\ & \times\times\times & & \times \\ & \times\times & \times & \times \\ & \times\times & & \times\times \\ 0 & 0 & \times\times\times & j \end{pmatrix}, \; j = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \text{ and } O = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \tag{5.6}$$

Here A' is an $n \times m$ matrix. The solution to this problem is

$$x^* = x + VA'(AVA')^{-1}(b - Ax). \tag{5.7}$$

If V is the identity matrix, we minimize the sum of the quadratic differences:

$$\sum_{j=1}^{n} (x_j^* - x_j)^2 . \tag{5.8}$$

A drawback of this function is the discontinuity that may arise between the last quarter of one year and the first quarter of the next (i.e. the step problem). In order to avoid this discontinuity, Denton considers the quadratic function based on the differences between the first-order differences, $\Delta x_j^* - \Delta x_j$. An attempt is made to preserve as much as possible the first-order differences of all successive quarters,

$$\sum_{j=1}^{n} (\Delta x_j^* - \Delta x_j)^2 \text{ with } \Delta x_j = x_j - x_{j-1}, \text{ and } \Delta x_1 = x_1 . \tag{5.9}$$

This quadratic function can be expressed in matrix form by choosing V as $V = D`D$, with

$$D = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix}, \text{ an } (n \times n) \text{ matrix.} \tag{5.10}$$

The Denton method preserves the initial value x_1 and all first order changes. The value of the first quarter x_1 is preserved, since Denton opts to work with an imaginary unmodifiable zero quarter, i.e. $x_0^* = x_0$, so that the first term in (5.9) equals $x_1^* - x_1$. A disadvantage of the above in practice is that it will often be undesirable to fix the level of x_1. To avoid this problem, Cholette (1984) uses a slightly adapted matrix

$$D = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & 1 & 0 \\ 0 & 0 & \cdots & 0 & -1 & 1 \end{pmatrix}. \tag{5.11}$$

Note that the number of rows is one fewer than the number of columns, requiring the calculation of a generalised inverse in (5.7).

5.4 Example

To illustrate the model, we created in Table 5.1 an artificial data set of twelve quarters and three annual totals. The quarterly data were chosen to include pronounced changes that follow the seasons. They must then be modified in a way that the sums of the four quarters for each year exactly equal the corresponding annual totals: 300, 400 and 500. We imposed no other constraints.

Table 5.1. Quarterly data before and after reconciliation.

	Original data	Reconciled data	Annual totals
Year 1	50	33	300
	100	73	
	150	120	
	100	74	
Year 2	50	36	400
	100	96	
	150	155	
	100	113	
Year 3	50	69	500
	100	124	
	150	178	
	100	129	

We applied the univariate Denton method with the Cholette matrix. It is simple here to calculate the reconciliation results from (5.7). We then rounded the results. The quarter-to-quarter changes were preserved as much as possible. The results of the first year were lower because of a lower annual total, and those of the third year were higher because of a higher annual total.

5.5 Characteristics

If no quarterly data are available, we could assume that all quarterly amounts are equal. Under this assumption the Denton method is comparable with the method proposed by Boot, Feibes and Lisman (1967).

Proportional differences are an alternative to arithmetic differences. They are applied with the idea that a large original value can be changed more in an absolute sense than a small original value. The reasoning is explained in section 6.3.

5.6 Quality indicators

As in the previous chapter, the most important quality indicator is *how* the high frequency series is modified. Of particular interest are the changes made in the first differences, given the starting point of the Denton method. The size of these changes is important, but the trend of the changes in time is particularly important. This trend can usually be assessed fastest by graphical means.

Another quality indicator of the Denton method is how accurately the high frequency series is aligned with the low frequency series. Numerical error will generally cause these differences to deviate slightly from zero. The differences are not usually a problem as long as they are less than a certain threshold value.

6. Multivariate Denton method

6.1 Short description

This section describes the extension to the Denton method for multiple variables.

Di Fonzo and Marini (2003 and 2005) combined the Denton method with Stone's (1942) method (see Chapter 3). Their method is used for rendering multivariate data coherent, and involves both constraints over time and constraints between the variables in one period. An example of a constraint in time is that the sum of four quarterly figures equals the corresponding annual figure (the so-called annual alignment). Annual alignment is normally achieved for each variable. Additional constraints cover other relationships in the data.

The aim of the Denton method is again to preserve changes in the high frequency series as much as possible, while satisfying all constraints.

Due to differences in the quality of the data it is desired that some variables will be adjusted more than others. Di Fonzo's and Marini's model is unable to handle these differences. Bikker and Buijtenhek (2006) further extended Di Fonzo's and Marini's model by adding reliability weights, allowing differences in reliability to be modelled. The method of Bikker and Buijtenhek is currently applied to the Dutch National Accounts.

6.2 Applicability

The conditions stated in section 5.2 also apply here.

The method as described is not intended for ratios, which can, however, be included in the model by applying a linearization technique (largely analogous to section 4.3.2).

6.3 Detailed description

This section provides a formal description of Di Fonzo's and Marini's reconciliation model. As in the previous section, we assume that quarterly data is to be aligned with annual data.

We use the same notation as in the previous sections. We denote the collection of quarterly data before reconciliation by x_{ij}, where i represents the various time series ($i = 1,\ldots,M$) and j represents the quarters ($j = 1,\ldots,n$). (In Chapter 5, $M = 1$.) The corresponding values after reconciliation are denoted by x_{ij}^*. We will represent the original and reconciled quarterly data as two ($Mn \times 1$) vectors x and x^*, which hold the data of all M variables for all n periods, viz.

$$x = (x_{11},\ldots, x_{1n}, x_{21},\ldots, x_{2n},\ldots., x_{M1},\ldots, x_{Mn})' \text{ and}$$

$$x^* = (x_{11}^*,\ldots, x_{1n}^*, x_{21}^*,\ldots, x_{2n}^*,\ldots., x_{M1}^*,\ldots, x_{Mn}^*)'. \qquad (6.1)$$

The constraints in this model are linear, and cover both reconciliation of the quarterly data to the annual data and relationships between different variables. The entire set of constraints is denoted as

$$Ax^* = b, \tag{6.2}$$

where A is a $(k \times Mn)$ matrix and b is a $(k \times 1)$ vector that also holds the annual figures, amongst others. The total number of constraints is k. Matrix A is an extension to matrix A in Equation (5.6). Here it holds that $k > n/4$ and the first $n/4$ rows of A determine the annual alignment, and therefore these rows are just the same as in A in (5.6). The last $k - n/4$ rows of A determine the constraints between the variables within a period or at different times.

As in Chapter 3, x^* is also the solution of the following quadratic optimization problem:

$$\min_{x^*} (x - x^*)' \, \Omega^{-1} (x - x^*), \tag{6.3}$$

$$\text{under the condition that } A\, x^* = b, \tag{6.4}$$

where Ω is the covariance matrix of x. Di Fonzo and Marini (2005) consider two forms for Ω. These two forms correspond with two models: an additive first difference (AFD) model and multiplicative, or proportional, first difference (PFD) model. The objective function in (6.3) of the additive model is

$$\text{AFD: } \sum_{i=1}^{M}\sum_{j=2}^{n} \frac{\left(\left(x_{ij}^* - x_{ij}\right) - \left(x_{ij-1}^* - x_{ij-1}\right)\right)^2}{v_{ij}} = \sum_{i=1}^{M}\sum_{j=2}^{n} \frac{1}{v_{ij}} \left(\Delta x_{ij}^* - \Delta x_{ij}\right)^2, \tag{6.5}$$

where $\Delta x_{ij}^* = x_{ij}^* - x_{ij-1}^*$, $\Delta x_{ij} = x_{ij} - x_{ij-1}$ and v_{ij} is a variance of the quarter-to-quarter changes in the numerator of (6.5). The objective function of the multiplicative model is:

$$\text{PFD: } \sum_{i=1}^{M}\sum_{j=2}^{n} \frac{1}{v_{ij}} \left(\frac{x_{ij}^*}{x_{ij-1}^*} - \frac{x_{ij}}{x_{ij-1}} \right)^2 = \sum_{i=1}^{M}\sum_{j=2}^{n} \frac{1}{v_{ij}} \left(\Gamma x_{ij}^* - \Gamma x_{ij}\right)^2, \tag{6.6}$$

where $\Gamma x_{ij}^* = x_{ij}^* / x_{ij-1}^*$ and $\Gamma x_{ij} = x_{ij} / x_{ij-1}$.

Bikker and Buijtenhek (2006) explain that the choice between an additive and a proportional objective function determines how the difference between the quarterly sum and the annual figure is dealt with.

An additive objective function is required if the quarter-to-quarter changes of the estimated variable must be as close as possible to the original changes, in absolute terms. The model assumption for this case is

$$\Delta x_{ij}^* - \Delta x_{ij} \sim \left(0, v_{ij}\right). \tag{6.7}$$

The additive model must also be used if we have no representative indicator for the change that we wish to measure, or the variable may be either positive or negative.

For the proportional model we assume that

$$I\tilde{x}_{ij}^{*} - I\tilde{x}_{ij} \sim \left(0, v_{ij}\right). \tag{6.8}$$

Which means that proportions between successive quarters must be preserved. This often occurs if the seasonal pattern has to be preserved.

We can restate (6.5) and (6.6) in matrix notation as

$$\text{AFD: } \Omega = \left(D'\left(X'V'VX\right)^{-1}D\right)^{-1}; \tag{6.9}$$

$$\text{PFD: } \Omega = X'\left(D'\left(V'V\right)^{-1}D\right)^{-1}X, \tag{6.10}$$

where D equals $I_M \otimes D_n$, and D_n is an $n \times n$ matrix

$$D_n = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & 1 & 0 \\ 0 & 0 & \cdots & 0 & -1 & 1 \end{pmatrix}. \tag{6.11}$$

In the previous section D_n was equal to D. Since we now have M variables, we also need M copies of the D_n matrix. This occurs through Kronecker multiplication. Again, all first-order changes and one level – the value of the first quarter of the first year – are preserved for each variable. Bikker and Buijtenhek (2006) consider alternatives for the D_n matrix.

Matrix X is diagonal, with the values from x on the diagonal. The V matrix has indicators on the diagonal that represent the reliability of the quarter-to-quarter changes relative to each other.

The great advantage of the multivariate Denton method is that it allows the additive and proportional models to be combined. This involves factorizing matrix X into two diagonal matrices X_{add} and X_{prop}, as follows: X_{add} has the same variables as X for the additive variables and a '1' for the proportional variables, whereas X_{prop} retains the original value for the proportional variables and a '1' for the additive variables (see Bikker and Buijtenhek (2006)). Matrix Ω can be rewritten as:

$$\Omega = X_{\text{prop}}\left(D'\left(X_{\text{add}}'V'VX_{\text{add}}\right)^{-1}D\right)^{-1}X_{\text{prop}}. \tag{6.12}$$

and if A is fully ranked, then (6.2) has no redundant constraints, and the solution for our optimization problem is

$$x^{*} = x + \Omega A'(A\Omega A')^{-1}(b - Ax). \tag{6.13}$$

An expression for the covariance matrix for x^{*} is

$$V^{*} = V - \Omega A'(A\Omega A')^{-1}AV. \tag{6.14}$$

6.4 Example

Table 6.1 has a data set with four variables and twelve quarters.

Table 6.1 Quarterly data before reconciliation

	Year 1				Year 2				Year 3			
	Q1	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**
Series												
x_1	335	399	335	351	355	364	312	366	335	364	335	351
x_2	347	379	343	365	341	371	333	342	336	377	389	381
x_3	340	365	338	356	333	332	351	356	340	365	338	356
x_4	341	371	337	359	335	361	337	350	350	370	348	200

The quarterly data must be modified such that the sum of four quarters exactly equals an annual total, as given in Table 6.2.

Table 6.2 Annual data

	Year 1	Year 2	Year 3
x_1	1350	1300	1350
x_2	1350	1300	1350
x_3	1350	1350	1400
x_4	1350	1350	1400

Besides the constraints in time, several additional constraints must be satisfied, i.e.

$$x_{1t} = x_{2t}, \quad t = 1,\ldots,12, \text{ and} \tag{6.15}$$

$$x_{3t} = x_{4t}, \quad t = 1,\ldots,12. \tag{6.16}$$

Note that the annual data satisfies (6.15) and (6.16), as indeed they must, since otherwise not all constraints can be fulfilled simultaneously.

We apply the multivariate Denton method and the proportional objective function is used for all variables. We assume that all diagonal entries of matrix V in (6.8) equal 0.2, which means that all quarter-to-quarter changes are equally reliable in relative terms. The results of the reconciliation are shown in Table 6.3.

Table 6.3 Reconciled data

	Year 1				Year 2				Year 3			
	Q1	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**
Series												
x_1	331	369	317	333	324	343	301	331	316	349	339	346
x_2	331	369	317	333	324	343	301	331	316	349	339	346
x_3	334	355	322	339	317	332	339	362	372	402	367	259
x_4	334	355	322	339	317	332	339	362	372	402	367	259

The quarterly figures of x_1 before and after reconciliation are shown in Figure 6.1.

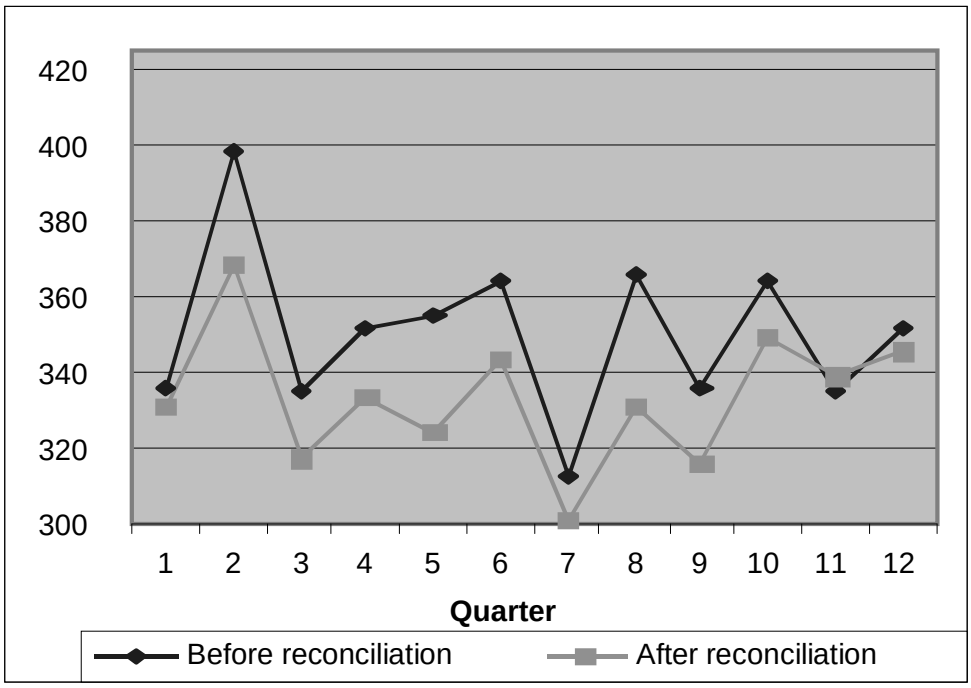

Figure 6.1. Time series of x_1.

The reconciliation process reduces the value of x_1 in all quarters except one. The reason is that the annual totals are less than the sum of the corresponding four quarterly figures. The original quarter-to-quarter changes have been preserved as much as possible. However, the change from the fourth to the fifth quarter is conspicuously different, in that before reconciliation it was positive, but after reconciliation it becomes negative. The reason is the large decrease in x_2 and the constraint that x_1 and x_2 must be equal.

6.5 Quality indicators

The most important quality indicator is again *how* the high frequency series are adjusted. Given the starting point of the Denton method, the changes in the first differences are of particular interest. The size of these changes is important, but the

trend of the changes in time is particularly important. This trend can usually be assessed fastest by graphical means.

The multivariate aspect demands that the changes of multiple series must be examined in their mutual context. The context is defined by the set of linear constraints. 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, as opposed to the adjustments.

Another important indicator is given by the ex post covariance matrix. The diagonal entries of this matrix provide information about the relative reliability of the reconciled results. Comparison with the covariance matrix used in the objective function yields information about how the reconciliation reduces the data variance. The nondiagonal entries of the ex post covariance matrix yield information about intervariable correlations introduced by reconciliation.

Another quality indicator of the multivariate Denton method is how accurately the high frequency series are aligned with the low frequency series. Numerical error will generally cause these differences to deviate slightly from zero. The differences are not usually a problem as long as they are less than a certain threshold value.

7. References

Bacharach, M. (1970), *Biproportional matrices & input-output change*. Cambridge University Press, Cambridge.

Bikker, R.P. and S. Buijtenhek (2006), *Alignment of Quarterly Sector Accounts to annual data*, Statistics Netherlands, Voorburg, http://www.cbs.nl/NR/rdonlyres/D918B487-45C7-4C3C-ACD0-oE1C86E6CAFA/0/Benchmarking_QSA.pdf.

Boonstra, H.J. (2006), *Macro-integratie Nationale Rekeningen, Een Bayesiaanse Benadering (Macro integration of National Accounts, a Bayesian Approach)*, Internal Report, Statistics Netherlands, Voorburg, http://cbsh1sps/sites/TmoMeth/analyse/07%20Analyse/2006/macrointegratienr.pdf.

Boot, J.C.G., W. Feibs, and J.H.C. Lisman (1967), *Further method of derivation of quarterly figures from annual data*. Applied Statistics **16** (1), 65-75.

Cholette, P.A. (1984), *Adjusting sub-annual series to yearly benchmarks*. Survey Methodology **10** (1), 35-49.

Chow, G.C. and A. Lin (1971), *Best Linear Unbiased Interpolation, and Extrapolation of Time Series by Related Series*. Rev. Economics and Statistics **53** (4), 372-375.

Dalen, J. Van (2002), *Winadjust: A program for adjusting matrices to given marginal totals*. Internal Report, Statistics Netherlands, Voorburg.

Denton, F.T. (1971), *Adjustment of Monthly to Quarterly Series to Annual Totals: An approach based on Quadratic Minimization*. Journal of the American Statistical Association **66** (333), 99-102.

Di Fonzo, T. and M. Marini (2003), *Benchmarking systems of seasonally adjusted time series according to Denton's moving preservation principle*. University of Padova, http://www.oecd.org/dataoecd/59/19/21778574.pdf.

Di Fonzo, T. and M. Marini (2005), *Benchmarking a system of Time Series: Denton's movement preservation principle vs. data based procedure*. University of Padova, http://epp.eurostat.cec.eu.int/cache/ITY_PUBLIC/KS-DT-05-008/EN/KS-DT-05-008-EN.pdf.

Harthoorn, R. and J. van Dalen (1987), *On the Adjustment of Tables With Lagrange Multipliers*, Occasional Paper NR/24, Statistics Netherlands, Voorburg.

Knottnerus, P. (2003), *Sample Survey Theory: Some Pythagorean perspectives*. Springer-Verlag, New York.

Magnus, J.R., J.W. van Tongeren, and A.F. de Vos (2000), *National Accounts Estimation using Indicator Ratios*, The Review of Income and Wealth **3**, 329-350, http://center.uvt.nl/staff/magnus/paper55.pdf.

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

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

Van Tongeren, J.W. (1986), *Development of an Algorithm for the Compilation of National Accounts and Related Systems of Statistics*. The Review of Income and Wealth **32**, 25-47.

Wroe D., P. Kenny, U. Rizki and I. Weerakoddy (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.

Version history

Version	Date	Description	Authors	Reviewers
Dutch version: Macro-integratie / Inpassen				
1.0	18-12-2007	First Dutch version	Reinier Bikker Jacco Daalmans Nino Mushkudiani	Piet Daas Eric Schulte Nordholt
1.1	23-01-2008	Small changes in the layout	Reinier Bikker Jacco Daalmans Nino Mushkudiani	
English version: Macro integration / Data reconciliation				
1.1E	17-02-2011	First English version	Reinier Bikker Jacco Daalmans Nino Mushkudiani	