Optimal adjustments for inconsistency in imputed data

Jeroen Pannekoek and Li-Chun Zhang

Discussion paper (201219)

Statistics Netherlands

The Hague/Heerlen, 2012

Explanation of symbols

.	data not available
*	provisional figure
**	revised provisional figure (but not definite)
x	publication prohibited (confidential figure)
–	nil
–	(between two figures) inclusive
0 (0.0)	less than half of unit concerned
empty cell	not applicable
2011–2012	2011 to 2012 inclusive
2011/2012	average for 2011 up to and including 2012
2011/'12	crop year, financial year, school year etc. beginning in 2011 and ending in 2012
2009/'10– 2011/'12	crop year, financial year, etc. 2009/'10 to 2011/'12 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

ISSN: 1572-0314

60083201219 X-10

Optimal adjustments for inconsistency in imputed data

Jeroen Pannekoek and Li-Chun Zhang [1]

Summary: Conflicting information may arise in statistical micro data due to partial imputation, where one part of the imputed record consists of the observed values of the original record and the other of the imputed values. Edit rules that involve variables from both parts of the record will often be violated. One strategy to remedy this problem is to make adjustments to the imputations such that all constraints are simultaneously satisfied and the adjustments are, in some sense, as small as possible. The minimal adjustments are obtained by minimizing a chosen distance metric subject to the constraints and we show how different choices of the distance metric result in different adjustments to the imputed data. As an extension we also consider an approach that does not aim to minimize the adjustments but to make the adjustments as uniform as possible between variables. Under this approach, even the values that are not explicitly involved in any constraints can be adjusted. The properties and interpretations of the proposed methods are illustrated using empirical business-economic data.

Keywords: Imputation, Edit-rules, consistent micro-data, optimization

[1]Li-Chun Zhang, Statistics Norway.

The authors are grateful for the many useful comments of Ton de Waal and Robbert Renssen on a previous draft of this paper.

1 Introduction

We are concerned with the task of reconciling conflicting information in statistical micro data that may arise due to partial (donor) imputation. The missing values are imputed either by the corresponding values of a suitable donor or by statistical estimation. The imputed record then contains two parts of data from different sources. One part contains the observed values from the original record and the other the imputed values. Edit rules that involve variables from both parts of the record will often be violated. For instance in business statistics we may have that *turnover* must be equal to the sum of *profit* and *costs*, where *costs* is again the sum of costs for material, personnel, housing *etc.*, and all the variables except profit must be non-negative. If some of the variables are missing, the imputed values taken from a donor in general will not automatically satisfy the various restrictions, together with the observed values of the original record. A numerical illustration based on structural business survey data will be given in Section 2.

Our strategy to remedy the inconsistency problem is to make adjustments to the imputed values that are minimal in some sense, such that a record consistent with the edit rules results. The edit rules are to be specified as linear equality-/inequality-constraints on the variables. The minimal adjustments are then obtained by minimizing a chosen distance metric subjected to these constraints. Using this optimization approach one is able to handle *all* the constraints at the same time. In comparison, traditional adjustment methods, such as prorating (e.g. Banff Support Team, 2008), only deal with one constraint at a time, which can be cumbersome and arbitrary whenever some variables are involved in multiple constraints. We will distinguish generally between *adjustable* (or *free*) variables that are allowed to be changed and *unadjustable* (or *fixed*) variables that are not to be changed. The distinction between free and fixed variables may coincide with the distinction between imputed and observed variables but this need not always be the case. For instance, some imputed values may be held fixed because they are derived by logical reasoning as in deductive imputation, or these may be obtained from external sources that are considered more reliable (or more suitable). On the other hand, there are cases where some observed values may be considered unreliable and are allowed to be changed. This optimization approach will be outlined in Section 2. In Section 3 we develop different ways to implement this optimization approach and in Section 4 we illustrate the properties of the proposed methods by an example. Finally, a summary and conclusions are provided in Section 5.

2 Outline of the optimization approach

2.1 Imputation of a business record with missing data

To illustrate the problem, we consider a small part of a record from a structural business survey with missing data that is to be imputed. The data for this record are shown in Table 1. Two response patterns are postulated; one with only Turnover observed and one where also Employees and Wages are observed. There are a number of common ways to impute the missing values in such a record. One possibility is the use of the values from a donor record to impute the missing values in the recipient record. This donor can, for instance, be the "nearest neighbour" donor record, from the same category of economic activity and closest to the recipient record in some metric based on some common observed variables, for instance Turnover for response pattern (I) and Employees, Turnover and Wages for response pattern (II). Imputation then entails the replacement of the missing values by the corresponding values from the donor record, we call this partial donor imputation because not all the values of the donor are transferred to the receptor.

Table 1. Data, missing data and donor values for variables in a business record. Explanation of abbreviated variable names: Employees (Number of employees); Turnover main (Turnover main activity); Turnover other (Turnover other activities); Turnover (Total turnover); Wages (Costs of wages and salaries).

Variable	*Name*	*Response (I)*	*Response (II)*	*Donor Values*
x_1	Profit			330
x_2	Employees		25	20
x_3	Turnover Main			1000
x_4	Turnover Other			30
x_5	Turnover	950	950	1030
x_6	Wages		550	500
x_7	Other Costs			200
x_8	Total Costs			700

2.2 The micro-level consistency problem

2.2.1 Introduction to the problem and some traditional solutions

Business records generally have to adhere to a number of accounting and logical constraints. These constraints are widely employed for checking the validity of a record and are, in this context, referred to as edit-rules. For the example record above, the following three edit-rules are formulated:

a1: $x_1 - x_5 + x_8 = 0$ (Profit = Turnover – Total Costs)

a2: $x_5 - x_3 - x_4 = 0$ (Turnover = Turnover main + Turnover other)

a3: $x_8 - x_6 - x_7 = 0$ (Total Costs = Wages + Other costs)

Partial donor imputation for either response pattern in Table 1 leads to violation of these edit-rules, which we refer to as the *(micro-level) consistency problem*. In particular, for response pattern (I), the first two edit-rules involving Turnover are violated and, for response pattern (II), all three edit-rules are violated. To obtain a consistent record some of the values have to be changed or "adjusted". Often, the imputed values are the candidates for adjustment while the actually observed values are not changed. However, other choices of adjustable and non-adjustable values can be made.

Traditional adjustment methods, such as the prorating method implemented in Banff (Banff Support Team, 2008), are designed to handle one constraint at a time. In response pattern (I), the prorating method could proceed as follows: (1) adjust the imputed values for Total costs and Profit with a factor 950/1030 to make them add up to the observed Turnover, (2) then adjust the imputed values for Turnover main and Turnover other with the same factor to satisfy the second edit and (3) adjust the imputed values of Wages and Other costs, also with the same factor to make them add up to the previously adjusted value of Total costs. Indeed, one may be tempted to extend this rescaling to imputed variables that are not in edit-constraints (only Employees in this case), which is not necessary for consistency with the specified edit-rules but can be justifiable if it is assumed that these variables are related to Turnover in approximately the same way as in the donor record. This last option is further discussed in Section 3.2.1.

This easy and intuitive solution becomes more complicated for the response pattern (II). Whereas the first two steps may be carried out as before, the third step shows some difficulties of this approach. Total costs appears in two edit-rules: *a2* and *a3*. In both edit-rules one variable is observed (Turnover and Wages, respectively) but Total costs is only adjusted to satisfy *a1* and the resulting adjusted value is irrespective of the observed value of Wages, thereby ignoring relevant information on the Total costs. Indeed, depending on the values available it can even happen that Total costs is adjusted downwards to the extend that it becomes smaller than Wages and hence there is no acceptable non-negative solution for Other costs. In general, adjusting a variable that appears in multiple edit-rules to just one of them is not only suboptimal in the sense described above, it also leads to rather arbitrary choices of the order in which edit-rules should be handled.

Another problem that the second response pattern illustrates is that a simple proportional adjustment is more plausible when variables have to be adjusted such that their *sum* equals a constant than when variables have to be adjusted in order to render their *difference* equal to a constant. For instance, for edit rule *a3*, formulated as Wages = Total costs − Other costs, with values $550 \neq$

700 − 200 and 550 fixed, a proportional adjustment of Total costs and Other costs would result in values of 770 and 220. However, much smaller adjustments can be obtained, e.g. by *increasing* Total costs to 740 while *decreasing* Other costs to 190. Such cases are therefore mostly excluded from prorating schemes.

2.2.2 A formal presentation of the edit constraints

For the further analysis of edit rules and adjustment methods it is convenient to express the edit restrictions in matrix notation, as $\mathbf{Cx} = \mathbf{b}$, where $\mathbf{C}$ is the *constraint* (or *restriction*) matrix and $\mathbf{b}$ is a constant vector. For the restrictions a1 – a3, $\mathbf{b} = \mathbf{0}$ and the matrix $\mathbf{C}$ is given by

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & -1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & -1 & 1 \end{pmatrix}$$

Notice that the non-zero elements in a row of the constraint matrix identify all the variables that are involved in the corresponding edit constraint, and the non-zero elements in a column of the constraint matrix identify all the edit constraints that involve the corresponding variable.

Successive moves from one non-zero element to another either in the same row or in the same column of a constraint matrix generates a *path*. A set of variables are *connected* (to each other) if there is a path between any two of them. A variable that is not connected with any other variables is an *isolated* variable. In the constraint matrix $\mathbf{C}$ above, the variable Employees (x_2) is an isolated variable, and the rest of the variables are connected. An example of a path between Profit (x_1) and Turnover main (x_3) is ($x_1 \rightarrow x_5 \downarrow x_5 \leftarrow x_3$), where the arrows indicate the direction of the successive movements along the path. Given a set of connected variables, a *joint* among them is a variable that has more than one non-zero element in the corresponding column of the constraint matrix. Different constraints are connected to each other through the joints. Indeed, two subsets of variables are *separated* from each other by a set of joints if any path between two variables, i.e. one from each subset, must pass through the set of joints. The joints of the matrix $\mathbf{C}$ here are (x_5, x_8). Moreover, (x_3, x_4) are separated from all the other variables by x_5, and (x_6, x_7) are separated from the others by x_8, and x_1 by (x_5, x_8).

From the constraint matrix, the following properties of the adjustment problem can be deduced:

i. An isolated variable, such as Employees (x_2) in Table 1, can be imputed (or adjusted) freely without causing consistency problems.

ii. Provided *all* the joints are observed or given by external sources, such as (x_5, x_8) in Table 1, the consistency problem among the set of connected variables can be resolved by dealing with one constraint at a time, e.g. by separate prorating for each constraint.

iii. Adjustments of a subset of variables do not cause consistency problems for the remaining connected variables given the joints that separate these variables. In Table 1, for instance, (x_3, x_4) can be adjusted freely given x_5 without causing consistency problem for the other variables.

iv. The imputation (or adjustment) of any variable may potentially cause consistency problems for all the connected variables that are not separated by the given joints. In both response patterns of Table 1 the joint x_5 is given. However, only (x_3, x_4) are separated from the other variables by x_5. The consistency problem among the rest of the connected variables (x_1, x_6, x_7, x_8) can be resolved using a traditional method in two steps: first, adjust the remaining joints (i.e. x_8) in a consistent manner given the observed joints (i.e. x_5); next, consider x_8 to be fixed and adjust the rest of the variables as in situation (***i***) and (***ii***). Thus, for response pattern (I), one might first impute x_8, say, proportionally to x_5. The remaining variables can be adjusted with regard to one-constraint at a time. For the response pattern (II), however, x_6 is also observed, such that it no longer seems desirable if one is to impute x_8 *without* taking into account x_6, because the two are connected. There arises therefore a need to deal with all the constraints that are connected to x_8 simultaneously, which requires an approach beyond the realm of traditional single-constraint adjustment methods such as prorating.

v. Constraints for which it is optimal to adjust the variables in the same direction (either an increase or a decrease) can be identified from the restriction matrix as rows in which the entries corresponding to adjustable variables have the same sign.

2.3 The optimization approach

One possible strategy to resolve the micro-level consistency problem introduced in section 2.2.1 is to adjust the imputed values, simultaneously and as little as possible, such that all edit-rules are satisfied. Denote by $\mathbf{x}_0$ the *adjustable* part of the record *before* adjustment and denote by $\tilde{\mathbf{x}}$ the corresponding sub-record *after* the adjustment, the optimization approach to the adjustment problem with *equality* constraints can be formulated as:

$$\begin{aligned} &\tilde{\mathbf{x}} = \arg\min_{\mathbf{x}} D(\mathbf{x}, \mathbf{x}_0) \\ &s.t. \quad \mathbf{A}\tilde{\mathbf{x}} = \mathbf{b}, \end{aligned} \tag{1}$$

with $D(\mathbf{x}, \mathbf{x}_0)$ a function measuring the distance (or deviance) between $\mathbf{x}$ and $\mathbf{x}_0$. In Section 3 we will consider different functions D for this adjustment problem. It is assumed here that there are no contradicting constraints, so that (1) can be solved (is feasible). Furthermore, for the algorithms we apply to solve (1) we will need to assume that $\mathbf{A}$ is of full row-rank, which means that, if necessary, redundant constraints have been removed from $\mathbf{A}$. Checks

for feasibility and redundancy of systems of edit rules can be performed with the R-package editrules (De Jonge and Van der Loo, 2011).

Some explanations on the notation in (1) are needed. To allow for the distinction between adjustable and non-adjustable variables we introduce the *accounting* matrix $\mathbf{A}$. If all the variables are adjustable, then $\mathbf{A} = \mathbf{C}$. Otherwise, the accounting matrix differs from the constraint matrix. Often, we will consider to adjust the imputed values only and leave the observed values unchanged, which means that the minimization in (1) is over the imputed values only. The complete data record may be partitioned into $\mathbf{x}_{obs}$ for the observed (and fixed) values and $\mathbf{x}_{mis}$ for the missing (and free) ones. A corresponding partition of the columns of the constraint matrix yields, say, $\mathbf{C}_{obs}$ and $\mathbf{C}_{mis}$. We have the following notational correspondence:

$$\mathbf{C}_{mis}\tilde{\mathbf{x}}_{mis} = \mathbf{A}\tilde{\mathbf{x}} = \mathbf{b} = -\mathbf{C}_{mis}\tilde{\mathbf{x}}_{obs} = -\mathbf{C}_{mis}\mathbf{x}_{obs}$$

More generally, we can distinguish between the adjustable and non-adjustable sub-records, which may not coincide with the distinction between the missing and observed sub-records, and write similarly, say,

$$\mathbf{C}_{free}\tilde{\mathbf{x}}_{free} = \mathbf{A}\tilde{\mathbf{x}} = \mathbf{b} = -\mathbf{C}_{fix}\tilde{\mathbf{x}}_{fix} = -\mathbf{C}_{fix}\mathbf{x}_{fix}$$

Clearly, the notation employed in (1) is much simpler, under the convention that the accounting matrix $\mathbf{A}$ always applies to the relevant adjustable variables, and so is any data vector (such as $\mathbf{x}$, $\tilde{\mathbf{x}}$ or $\mathbf{x}_0$) that it operates on. Notice that, while the constraint matrix $\mathbf{C}$ is derived *a priori* from the edit-rules alone, without reference to the actual data, and is the same for all the records, the accounting matrix $\mathbf{A}$ is generally different from one record to another, such as when the adjustable values are simply the imputed values.

Now, in addition to the equality constraints, we often have linear inequality constraints. The simplest case is the non-negativity of most economic variables. Other inequality constraints may arise, for instance, when it is known (or required) that *Wages* should not be less than a certain factor, say, $f_{\min}$ (i.e. the minimum wage) times *Employees*. The optimization problem can be extended to handle inequality constraints by formulating the constraints as

$$\begin{aligned} \mathbf{A}_{eq}\tilde{\mathbf{x}} &= \mathbf{b}_{eq} \\ \mathbf{A}_{ineq}\tilde{\mathbf{x}} &< \mathbf{b}_{ineq} \end{aligned} \tag{2}$$

where $\mathbf{A}_{eq}$ contains the rows of $\mathbf{A}$ corresponding to equality constraints and $\mathbf{A}_{ineq}$ the ones corresponding to inequality constraints. In either case, both the contributions of the non-adjustable variables and all the possible constants involved are combined into the $\mathbf{b}$-vectors. For ease of exposition we shall, unless noting otherwise, write these equality/inequality constraints more compactly as $\mathbf{A}\tilde{\mathbf{x}} \leq \mathbf{b}$.

3 Development of the optimization approach

In this section we will develop different ways to implement an optimization approach to the consistency problem and show the different properties, purposes and assumptions of the resulting adjustment methods. In section 3.1 methods are discussed that aim at adjusting imputed values to attain consistency with the edit rules and keeping these adjustments as small as possible. In section 3.2.1 a method is discussed that also adjusts imputed values to attain consistency but assumes that these adjustments should be as uniform as possible rather than as small as possible.

3.1 Minimum adjustments: three distance functions and corresponding adjustment models

The conditions for a solution to the minimization problem formulated in (1) with constraints (2) can be found by inspection of the Lagrangian for this problem, which can be written as

$$\begin{aligned} L(\mathbf{x}, \boldsymbol{\alpha}) &= D(\mathbf{x}, \mathbf{x}_0) + \boldsymbol{\alpha}^T(\mathbf{A}\mathbf{x} - \mathbf{b}) \\ &= D(\mathbf{x}, \mathbf{x}_0) + \boldsymbol{\alpha}_{eq}^T(\mathbf{A}_{eq}\mathbf{x} - \mathbf{b}_{eq}) + \boldsymbol{\alpha}_{ineq}^T(\mathbf{A}_{ineq}\mathbf{x} - \mathbf{b}_{ineq}) \\ &= D(\mathbf{x}, \mathbf{x}_0) + \sum_{k \in I_{eq}} \alpha_k(\mathbf{a}_k^T\mathbf{x} - b_k) + \sum_{k \in I_{ineq}} \alpha_k(\mathbf{a}_k^T\mathbf{x} - b_k), \end{aligned} \tag{3}$$

with $\boldsymbol{\alpha}$ a vector of Lagrange multipliers, or *dual* variables, with components α_k, one for each of the K constraints, and $\mathbf{a}_k$ the k-th row (corresponding to constraint k) of the $K \times J$ accounting matrix $\mathbf{A}$, with J the number adjustable variables which equals the number of elements of $\mathbf{x}$. In accordance with the partitioning of $\mathbf{A}$ and $\mathbf{b}$ corresponding to the type (equality/inequality) of the constraints, the dual vector can also be partitioned into a subvector $\boldsymbol{\alpha}_{eq}$ pertaining to the equality constraints and a subvector $\boldsymbol{\alpha}_{ineq}$ pertaining to the inequality constraints. The distinction between equality and inequality constraints can also be made explicit by defining the index sets I_{eq} and I_{ineq} containing the indices of the equality and inequality constraints respectively. This is used in the last line of (3).

From optimization theory it is well known that for a convex function $D(\mathbf{x},\, \mathbf{x}_0)$ and linear (in)equality constraints, the solution to this constrained optimization problem is given by vectors $\tilde{\mathbf{x}}, \tilde{\boldsymbol{\alpha}}$ that satisfy the so-called Karush-Kuhn-Tucker (KKT) conditions (see, e.g. Luenberger, 1984, Boyd and Vandenberghe, 2004). One of these conditions is that the gradient of the Lagrangian w.r.t. $\mathbf{x}$ is zero when evaluated at the optimal point $\tilde{\mathbf{x}}, \tilde{\boldsymbol{\alpha}}$, i.e.

$$L'_{x_j}(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\alpha}}) = D'_{x_j}(\tilde{\mathbf{x}}, \mathbf{x}_0) + \sum_k \tilde{\alpha}_k a_{ki} = 0, \text{ for all } x_j, \tag{4}$$

with $L'_{x_j}(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\alpha}})$ being the gradient of L w.r.t. x_j evaluated at $\mathbf{x} = \tilde{\mathbf{x}}$ and $\boldsymbol{\alpha} = \tilde{\boldsymbol{\alpha}}$, and D'_{x_j} that of D. From this condition alone, we can already see how different

choices for D lead to different solutions to the adjustment problem. In the next three subsections we shall consider three familiar choices for D: Least Squares, Weighted Least Squares and Kullback-Leibler divergence, and show how these different choices result in different structures of the adjustments, which we will refer to as the *adjustment models*.

In addition to (4), and apart from $\tilde{\mathbf{x}}$ satisfying the constraints, the other KKT conditions prescribe that for inequality constraints we must also have

$$\tilde{\boldsymbol{\alpha}}_{ineq} \geq \mathbf{0}; \tag{5}$$

$$\tilde{\alpha}_{ineq,k}(\mathbf{a}_k^T\tilde{\mathbf{x}} - b_k) = 0 \quad k \in I_{ineq} \tag{6}$$

with $\alpha_{ineq,k}$ an element of $\boldsymbol{\alpha}_{ineq}$. Condition (5) is component-wise non-negativity, which is only required for $\tilde{\boldsymbol{\alpha}}_{ineq}$ and not for $\tilde{\boldsymbol{\alpha}}_{eq}$. Condition (6) is termed "complementary slackness" and states that if an inequality constraint is satisfied with "slack", that is the residual of the constraint: $\mathbf{a}_k^T\tilde{\mathbf{x}} - b_k$ is *strictly* less than zero, then the corresponding dual variable is zero and, conversely, if the dual variable is positive then the residual is zero and the constraint is satisfied with equality. These conditions will provide additional insight in the properties of the adjusted vector $\tilde{\mathbf{x}}$ and in the algorithm, described in the appendix, to arrive at the solution to this optimization problem.

3.1.1 Least Squares

First, we consider the least squares (LS) criterion to find an adjusted x-vector that is closest to the original unadjusted data, that is: $D(\mathbf{x}, \mathbf{x}_0) = \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^T(\mathbf{x} - \mathbf{x}_0)$, and so $D'_{x_j}(\tilde{\mathbf{x}}, \mathbf{x}_0) = \tilde{x}_j - x_{0,j}$. We then obtain from (4)

$$\tilde{x}_j = x_{0,j} - \sum_k a_{kj}\tilde{\alpha}_k. \tag{7}$$

This shows that the LS-criterion results in additive adjustments: the total adjustment to variable $x_{0,j}$ is the sum of adjustments to each of the constraints k. These adjustments consist of adjustment parameters (the dual variables) $\tilde{\alpha_k}$ that describe the amount of adjustment due to constraint k and variables a_{kj} (the elements of the accounting matrix) describing the specific adjustments to the variables $x_{0,j}$. Often (but not always) the a_{kj} are 1, -1 or 0 and the corresponding $x_{0,j}$ is adjusted by, respectively, $\tilde{\alpha}_k$, $-\tilde{\alpha}_k$ or not at all. A smaller value for an $\tilde{\alpha}_k$ (in absolute value if $k \in I_{eq}$) corresponds to smaller adjustments and hence a more optimal value of the objective $D(\mathbf{x}, \mathbf{x}_0)$ to be minimized, all other $\tilde{\alpha}_k$ held fixed. A zero value for an $\tilde{\alpha}_k$ means that no adjustment to that specific constraint has taken place. For inequality constraints this means that the residual of the constraint is zero or negative. On the other hand if, for inequality constraints, $\tilde{\alpha}_k$ is positive, adjustment has taken place and the residual of the constraint is zero according to (6). Similar interpretations hold,

but will not be repeated, for the adjustment models presented in sections 3.1.2 and 3.1.3.

3.1.2 Weighted Least Squares

For the weighed least squares (WLS) criterion, $D(\mathbf{x}, \mathbf{x}_0) = \frac{1}{2}(\mathbf{x}-\mathbf{x}_0)^T Diag(\mathbf{w})(\mathbf{x}-\mathbf{x}_0)$, with $Diag(\mathbf{w})$ a diagonal matrix with a vector of weights along the diagonal. The derivative of this loss function is $w_j(\tilde{x}_j - x_{0,j})$ and we obtain from (4)

$$\tilde{x}_j = x_{0,j} - \frac{1}{w_j}\sum_k a_{kj}\alpha_k. \tag{8}$$

Contrary to the least squares case where the amount of adjustment to a constraint is equal in absolute value (if it is not zero) for all variables in that constraint, the amount of adjustment now varies between variables according to the weights: variables with large weights are adjusted less than variables with small weights. The weighted least squares approach to the adjustment problem has been applied by Thomson et al. (2005). They used weights of 10,000 for observed values and weights of 1 for imputed values. Effectively, this means that if a consistent solution can be obtained by changing only imputed variables, this solution will be found. Otherwise some observed variables will also be adjusted. This is an example of the distinction between "missing *vs.* observed" and "adjustable *vs.* fixed".

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

$$\frac{\tilde{x}_j}{x_{0,j}} = 1 - \sum_k a_{kj}\alpha_k, \tag{9}$$

which is an additive model for the *ratio* between the adjusted and unadjusted values. It may be noticed that this is the first-order Taylor expansion (i.e. around 0 for all the α_k's) to the multiplicative adjustment given by

$$\frac{\tilde{x}_j}{x_{0,j}} = \prod_k (1 - a_{kj}\alpha_k) \tag{10}$$

From (9) we see that the α_k's determine the difference from 1 of the *ratio* between the adjusted and pre-adjusted values, which is usually much smaller than unity in absolute value (e.g. an effect of 0.2 implies a 20% increase due to adjustment which is large in practice). The products of the α_k's are therefore often much smaller than the α_k's themselves, in which cases (9) becomes a good approximation to (10), and one may regard the WLS adjustment to be roughly given as the product of all the constraint-specific multiplicative adjustments.

3.1.3 Kullback-Leibler divergence

The Kullback-Leibler (KL) divergence measures the difference between $\mathbf{x}$ and $\mathbf{x}_0$ by the function $D_{KL} = \sum_j x_j(\ln x_j - \ln x_{0,j} - 1)$. Its derivative w.r.t. x_j and evaluated at $\tilde{\mathbf{x}}$ is $\ln \tilde{x}_j - \ln x_{0,j}$ and we obtain from (4) the following adjustment model

$$\tilde{x}_j = x_{0,j} \prod_k \exp(-a_{kj}\alpha_k). \tag{11}$$

In this case the adjustments have a multiplicative form and the adjustment for each variable is the product of adjustments due to each of the constraints. The adjustment due to constraint k is equal to 1 if a_{kj} is 0 (i.e. no adjustment), it is $1/\exp(\alpha_k)$ if a_{kj} is 1 and it is $\exp(\alpha_k)$ if a_{ik} is -1.

3.1.4 Explicit solution for weighted least squares without inequality constraints

If the loss function is weighted least squares and the constraints are only equality constraints, there is an explicit solution for the optimization problem that will be given below. For problems with inequality constraints and for the KL-divergence, iterative methods will be needed in general. An iterative method that is especially suited for the problems considered here is treated in the appendix.

For weighted least squares the Lagrangian is $L(\mathbf{x}, \boldsymbol{\alpha}) = \frac{1}{2}(\mathbf{x}-\mathbf{x}_0)^T Diag(\mathbf{w})(\mathbf{x}-\mathbf{x}_0) + \boldsymbol{\alpha}^T(\mathbf{A}\mathbf{x}-\mathbf{b})$, and the solution to the optimization problem with equality constraints only is the vector $\tilde{\mathbf{x}}$ that solves the equations

$$L'_{\mathbf{x}}(\mathbf{x}, \boldsymbol{\alpha}) = Diag(\mathbf{w})(\mathbf{x} - \mathbf{x}_0) + \mathbf{A}^T\boldsymbol{\alpha} = \mathbf{0} \tag{12}$$

$$L'_{\boldsymbol{\alpha}}(\mathbf{x}, \boldsymbol{\alpha}) = \mathbf{A}\mathbf{x} - \mathbf{b} = \mathbf{0}. \tag{13}$$

Solving (12) for $\mathbf{x}$ we obtain $\mathbf{x} = \mathbf{x_0} - Diag(\mathbf{w})^{-1}\mathbf{A}^T\boldsymbol{\alpha}$ and substituting this result in (13) yields

$$\tilde{\boldsymbol{\alpha}} = (\mathbf{A}Diag(\mathbf{w})^{-1}\mathbf{A}^T)^{-1}(\mathbf{A}\mathbf{x}_0 - \mathbf{b})$$

and hence, we have from (12)

$$\tilde{\mathbf{x}} = \mathbf{x}_0 - \mathbf{A}^T(\mathbf{A}Diag(\mathbf{w})^{-1}\mathbf{A}^T)^{-1}(\mathbf{A}\mathbf{x}_0 - \mathbf{b}). \tag{14}$$

3.2 Generalized ratio adjustments

3.2.1 The generalized ratio model

In section 2.2.1 we considered, for the response pattern I in which only Turnover was observed, a simple ratio adjustment which entailed the multiplication of all imputed values by the ratio between the observed Turnover value and the

corresponding donor value. This ratio adjustments results in a record that satisfies all constraints because the donor record did. It is, however, different from a minimal adjustment approach because more variables are adjusted than is necessary to satisfy the constraints. In particular, the value of Employees is modified even if it does appear in any of the constraints. The motivation for such a ratio-adjustment is to serve two purposes: adjustment to satisfy the constraints and improving on the imputed values by taking assumed relations between the variables into account. In this case it is assumed that all imputed variables are related to Turnover and that since the donor Turnover value is larger than the recipient value the imputed (donor) values can made to better "fit" the recipient record by scaling them all down with the factor (Turnover observed)/ (Turnover donor). As discussed in section 2.2.1 this simple approach cannot, in general, be applied without consistency problems if there is more than one variable observed and there are several ratio's between observed and donor variables to be considered simultaneously. In this section we develop a generalization of this ratio adjustment that can take multiple ratio's into account and will lead to a consistent record even if the donor values or otherwise imputed values are already violating edit-rules.

For what we call the *generalized ratio (GR) adjustments*, we consider the following adjustment model

$$\tilde{x}_j = x_{0,j}\delta_j$$

i.e. component-wise multiplicative adjustment. The δ_j will be set to 1 for fixed variables which will usually include at least the observed ones and will be unequal to 1 for all other variables including imputed variables that stand in constraints as well as imputed variables that are not part of any constraint. In concordance with the simple ratio adjustment we will try to find δ_j that are as uniform as possible and also result in a consistent adjusted $\tilde{\mathbf{x}}$-vector. This clearly entails the simple ratio adjustment (with constant δ_i) as a special case. To find optimal values for the δ_j we consider the following objective function

$$\Delta(\delta) = \frac{1}{2}\sum_{j=1}^{J'}(\delta_j - \bar{\delta})^2, \quad \text{where} \quad \bar{\delta} = \frac{1}{J'}\sum_j \delta_j \tag{15}$$

where δ is the vector collecting all δ_j's, $\bar{\delta} = \frac{1}{J'}\sum_j \delta_j$ and J' denotes all the variables under consideration, fixed or free, rather than just the J adjustable ones. The adjustment factors δ_j are obtained by minimizing (15) subject to the edit constraints. Since $\delta_i = \tilde{x}_i/x_{0,i}$ and the $x_{0,i}$ are fixed, this is a minimization problem of the form (1) with constraints (2), i.e.

$$\tilde{\mathbf{x}} = \arg\min_{\mathbf{x}} \Delta(\mathbf{x}, \mathbf{x}_0) \quad s.t. \quad \mathbf{A}\tilde{\mathbf{x}} \leq \mathbf{b}.$$

We consider the loss function Δ to aim at a kind of most-uniform adjustment solution that is a generalization of the simple ratio adjustments. For instance, for response pattern (II), where there are three observed values that do not

share a common ratio towards the corresponding donor values, the generalized ratio adjustments are given by the component-wise ratio's that deviate least from each other, while taking into account the three observed ones (with $\delta's$ equal to 1) as well as the edit-constraints.

Uniform adjustment of all adjustable variables is a statistical assumption or a 'soft' constraint that the adjusted record does not necessarily have to satisfy, in contrast to the 'hard' constraints expressed in the edit rules. A case in favour of the generalized ratio method can be made if the differences between the values in the donor and receptor records are assumed to be in the same direction and of approximately the same size. This could be the case for donor imputation in business statistics where all donor values are from the same donor record and the donor and receptor are businesses with similar structure but different size. The generalized ratio method seems to be less appropriate when the imputed values are from different donors or when the the imputed values are model based predictions because in such cases there is no reason that adjustments should be in the same direction or similar in size.

The distance metrics considered for the minimum adjustment approaches in the previous subsection can be characterized as *decomposable*, in the sense that the overall distance between two vectors is given as a (weighted) sum of 'distances' between the corresponding components. The loss-function for the generalized ratio adjustments is the empirical variance of the component-wise adjustments which is a *non-decomposable* loss function, where each adjustment is dependent on the other adjustments. A consequence is that a variable that is not part of any constraint will retain the initial (donor) value under the minimum adjustment approach, but using generalized ratio adjustments these variables will be adjusted because of the changes made to the variables that are constraint-bound.

In the special case where a unit has no observed value at all, the minimum adjustment approach would lead to imputation of all the donor values *without* any adjustment if they are consistent, whereas the most-uniform adjustments are not well defined because some constraints towards observed values are necessary in order to identify a unique solution.

3.2.2 Explicit solution for problems without inequality constraints

If the constraints are only equalities, an explicit solution for the δ_j will be given below. For the more general problem with inequalities we have no tailor made algorithm yet, but general optimization routines can be applied.

Let δ be partitioned in a part containing the fixed values and a part containing the free values so that we have $\delta = (\delta_{free}^T, \delta_{fixed}^T)^T$. Then we can write the Lagrangian for Δ in (15) as

$$L = \Delta(\delta) + \alpha^T(\mathbf{A}\mathbf{x} - \mathbf{b}) = \Delta(\delta) + \alpha^T\lambda \quad \text{where } \lambda = \mathbf{A}\mathbf{x} - \mathbf{b} = \mathbf{A}(\mathbf{x}_0 \times \delta_{free}) - \mathbf{b},$$

α is the vector of Lagrange multipliers, and $\mathbf{x}_0 \times \delta$ denotes element-by-element vector multiplication, i.e. $\mathbf{x}_0 \times \delta_{free} = (x_{0,1}\delta_{free,1}, ..., x_{0,J}\delta_{free,_J})^T$. Note that as before the accounting matrix $\mathbf{A}$, and the corresponding vectors $\mathbf{x}$, $\mathbf{x}_0$ and δ_{free} all pertain to the adjustable variables whereas the δ in $\Delta(\delta)$ includes all δ_j's, either fixed or free.

Now, define $\mathbf{P} = \mathbf{I}_{J\times J} - \mathbf{1}_{J\times 1}\mathbf{1}_{J\times 1}^T/J'$, where I is the identity matrix and $\mathbf{1}$ a vector with ones and let $\mathbf{W}$ be the $J \times (J' - J)$-matrix whose elements are all given by $-1/J'$. Then, we can express the derivative of L w.r.t. δ_{free} as

$$\partial L/\partial\delta_{free} = \partial\Delta/\partial\delta_{free} + \partial(\alpha^T\lambda)/\partial\delta_{free} = \mathbf{P}\delta_{free} + \mathbf{W}\delta_{fixed} + \mathbf{Z}^T\alpha, \quad (16)$$

where

$$\mathbf{Z}^T = \begin{pmatrix} a_{11}x_{0,1} & \cdots & a_{K1}x_{0,1} \\ a_{12}x_{0,2} & \cdots & a_{K2}x_{0,2} \\ \vdots & \cdots & \vdots \\ a_{1J}x_{0,J} & \cdots & a_{KJ}x_{0,J} \end{pmatrix}$$

The derivative of L w.r.t. the Lagrange multipliers α can be written as

$$\partial L/\partial\alpha = \partial(\alpha^T\lambda)/\partial\alpha = \mathbf{Z}\delta_{free} - \mathbf{b} \quad (17)$$

By setting (16) to zero we obtain

$$\delta_{free} = -\mathbf{P}^{-1}(\mathbf{W}\delta_{fixed} + \mathbf{Z}^T\alpha). \quad (18)$$

Substituting the expression (18) for δ_{free} into (17) and setting the resulting equation to zero yields first α and then δ_{free} as follows

$$\alpha = -(\mathbf{Z}\mathbf{P}^{-1}\mathbf{Z}^T)^{-1}(\mathbf{b} - \mathbf{Z}\mathbf{P}^{-1}\mathbf{W}\delta_{fixed})$$
$$\delta_{free} = -\mathbf{P}^{-1}\mathbf{W}\delta_{fixed} + \mathbf{P}^{-1}\mathbf{Z}^T(\mathbf{Z}\mathbf{P}^{-1}\mathbf{Z}^T)^{-1}(\mathbf{b} - \mathbf{Z}\mathbf{P}^{-1}\mathbf{W}\delta_{fixed})$$

4 Example revisited

All the adjustments methods (LS, WLS, KL and GR) have been applied to the example record in Table 1. For the WLS method we used as weights the inverse of $\mathbf{x}_0$ so that the relative differences between $\mathbf{x}$ and $\mathbf{x}_0$ are minimized and the adjustments are proportional to the components of $\mathbf{x}_0$. For this choice of weights, the KL- and WLS-methods lead to results that are equal up to the first decimal. The results for both response patterns are given in Table 2. The observed values are treated as fixed and shown in bold, the imputed values are adjustable.

For both response patterns, the LS adjustment procedure leads to one negative value (for *Turnover other*) which is not acceptable (Table 2). Therefore the LS-procedure was run again with a non-negativity constraint added for the variable *Turnover other*. This results simply in a zero for that variable and a

value of 950 for *Turnover main* to ensure that *Turnover = Turnover main + Turnover other*. Without the non-negativity constraint, the LS-results clearly show that for variables that are part of the same constraints (in this case the pairs of variables x_3, x_4 and x_6, x_7 that are both appearing in one constraint only), the adjustments are equal: -40 for x_3, x_4 and -16 for x_6, x_7. *Total costs* (x_8) is part of two constraints and therefore the total adjustment to this variable consists of two additive components. One component to adjust to the constraint *a1*:x_1- x_5+ x_8= 0 (*Profit = Turnover – Total Costs*) and one component to adjust to *a3*: x_8- x_6- x_7= 0 (*Total Costs = Wages + Other costs*). For response pattern (I), the first component is minus 48 - which is also the single adjustment component for *Profit* - and the second component is 16 – which is also the single adjustment component for *Wages* and *Other costs* (with opposite sign). These two components add up to the adjustment of -32.

The results for the WLS/KL solution show that for this weighting scheme the adjustments are larger, in absolute value, for large values of the imputed variables than for smaller ones. In particular, the adjustment to *Turnover other* is only -2.3 - so that no negative adjusted value results in this case - whereas the adjustment to *Turnover main* is 77.7. The multiplicative nature of these adjustments (as KL-type adjustments) also clearly shows since the adjustment *factor* for both these variables is 0.92 (for both response patterns). The adjustment factor for *Wages* and *Other costs* in response pattern (I) is also equal (to 0.94) because these variables are in the same single constraint and so the ratio between these variables is unaffected by this adjustment. However the ratio of each of these variables to *Total Costs* is not unaffected because *Total Costs* has a different sign in the constraint *a3* and, moreover, *Total Costs* is also part of constraint *a1* so that it is subject to two adjustment factors.

Table 2. Imputation and adjustment of business record in Table 1. DI: Direct partial donor imputation without adjustment; LS: Minimum Least-squares adjustments; WLS: Minimum weighted least-squares adjustments; KL: Minimum Kullback-Leibler divergence adjustments; GR: Generalized ratio adjustments.

	Response (I)				*Response (II)*			
Variable	DI	LS	WLS/KL	GR	DI	LS	WLS/KL	GR
x_1	330	282	291	304	330	260	249	239
x_2	20	20	20	18	**25**	**25**	**25**	**25**
x_3	1000	960	922	922	1000	960	922	921
x_4	30	-10	28	28	30	-10	28	29
x_5	**950**	**950**	**950**	**950**	**950**	**950**	**950**	**950**
x_6	500	484	470	461	**550**	**550**	**550**	**550**
x_7	200	184	188	184	200	140	151	161
x_8	700	668	658	646	700	690	701	711

As expected, the generalized ratio adjustments reduce to a global proportional adjustment of all the imputed values by a ratio of 0.922 (=950/1030) for re-

sponse pattern (I), including the variable Employee. This is a main difference from the minimum-adjustment methods that are based on decomposable loss functions. For response pattern (II), the GR adjustments are closer to the WLS/KL solution than to the LS solution.

Table 3. Criterion values for the three adjustment methods.

Method	*Loss function value*		
	LS	WLS	GR
LS	20925	78.0	0.1434
WLS/KL	23976	50.6	0.0276
GR	25090	51.6	0.0270

In table 3 we listed the values of the loss functions for the solutions of the different adjustment methods. For the (weighted) least squares loss function we actually took 2 times the loss function value such that it equals the (weighted) sum of squares. For each column in table 3, the smallest values are in the diagonal cells, meaning that each method minimizes its corresponding loss function, as it should be. For all three loss functions, it appears that the differences between the generalized ratio adjustments and the WLS/KL solution are smaller than the differences between those two solutions and the LS solution. For instance, the empirical variance of the multiplicative factors (i.e. the loss function Δ) is 0.0270 by the GR adjustments and 0.0276 for the WLS/KL solution, but is increased to 0.1434 for the LS solution. A similar large increase in the loss function value of the LS solution occurs for the WLS loss function. For the LS loss function, the differences between the three methods are not so pronounced.

5 Discussion

Imputation is generally used as a method to compensate for partially missing values. Often, especially in structural business data, the data have to satisfy many carefully specified edit-rules, derived from logical relations or accounting equations. Traditional approaches to imputation can not handle such edit-rules at the same time and, as a consequence, inconsistencies will arise in the imputed micro-data. Ad-hoc post-imputation adjustments, often applied in a somewhat arbitrary sequence, are not only undesirable in theory, but can also be tedious to implement. Using the edit-rules to adjust the imputed records such that they simultaneously conform to all edit-rules is a more satisfactory approach.

In this paper we have formulated an optimization approach to solve the simultaneous adjustment problem. Two variations of the optimization criterion have been considered. The first one seeks to minimize the adjustments needed to ensure consistency. In this approach only variables that appear in edit-rules will be adjusted because other variables will not cause inconsistency problems. The second is called the "generalized ratio" approach. In this approach all imputed values are adjusted and the adjustments are as uniform across variables

as possible. The inconsistency is seen as an indication that there are systematic differences between the donor values and the observed values and it may therefore be plausible to adjust all imputed variables.

The optimization approach to the inconsistency problem provides a general methodology that extends beyond the traditional single-constraint adjustment methods such as prorating. All constraints are handled simultaneously and, if variables appear in more than one constraint then they are adjusted according to all of them. Besides being an optimal method according to the chosen distance metric or loss function, this simultaneous approach also has the practical advantage that there is no need to specify the order in which the constraints are to be applied.

For the minimum adjustment approach several distance metrics have been analysed. It is shown that (weighted) least-squares loss function leads to additive adjustments and that minimizing the Kullback-Leibler information criterion leads to multiplicative adjustments. It is also shown that for a specific choice of weights the WLS solution is an approximation to the KL solution.

When the statistical assumptions underlying the generalized ratio method are met, we expect similar results of multiplicative adjustments by the minimum adjustment approach and by the generalized ratio method as far as the variables that appear in the constrained are considered. However, the GR method also adjusts other imputed variables that are not in the constraints, whereas the minimimal adjustment methods leave these variables unchanged. Adjusting all imputed values can be motivated, in the case of donor imputation, by assuming that the differences between the values in the donor and receptor records are in the same direction and of approximately the same size. This could be the case for donor imputation in business statistics where all donor values are from the same donor record and the donor and receptor are businesses with similar structure but different size. The generalized ratio method seems to be less appropriate when the imputed values are from different donors or when the imputed values are model based predictions because in such cases there is no reason that adjustments should be in the same direction or similar in size.

References

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

Boyd, S., and L. Vandenberghe (2004). *Convex Optimization*, Cambridge University Press, Cambridge.

Cenzor, Y., and S.A. Zenios (1977). *Parallel Optimization. Theory, Algorithms, and Applications.* Oxford University Press, New York.

De Jonge, E. and M. Van der Loo (2011). *Manipulation of linear edits and er-*

ror localization with the editrules package. Technical Report 201120, Statistics Netherlands, The Hague.

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

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

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

Appendix A. The successive projection algorithm

In this appendix we briefly review an algorithm that can be used to solve the optimization problem that was formulated in section 2.3 as:

$$\begin{aligned} &\tilde{\mathbf{x}} = \arg\min_{\mathbf{x}} D(\mathbf{x}, \mathbf{x}_0) \\ &s.t. \quad \mathbf{A}\tilde{\mathbf{x}} \leq \mathbf{b}, \end{aligned} \tag{19}$$

with $D(\mathbf{x}, \mathbf{x}_0)$ a convex function measuring the distance (or deviance) between $\mathbf{x}$ and $\mathbf{x}_0$ and $\mathbf{A}\tilde{\mathbf{x}} \leq \mathbf{b}$ representing linear equality and inequality constraints. This convex optimization problem was solved explicitly in section 3.1.4 if the objective function is the (weighted) least squares function and there are only equality constraints. For more general problems when D is any convex function and there are also inequality constraints several optimization methods can be used. In the remainder of this appendix we will give some details of applying a general iterative approach, referred to as the Successive Projection Algorithm (SPA), to the distance functions considered in this paper. This algorithm is easy to implement and contains as a special case the – among survey methodologists well known – Iterative Proportional Fitting (IPF) algorithm (also known as Raking) for adjusting contingency tables to new margins. Algorithms of this type are extensively discussed in Censor and Zenios (1997) and some applications to adjustment problems are described in De Waal et al. (2011).

For the convex optimization problems considered here, we need not solve the *primal* problem (19) directly, we can also solve the primal problem by solving the *dual* problem first, which in our applications turns out to lead to a particularly simple algorithm. The dual function associated with the Lagrangian $L(\mathbf{x}, \boldsymbol{\alpha})$ for the problem (19) is obtained by minimizing the Lagrangian w.r.t. $\mathbf{x}$ and substituting the resulting value for $\mathbf{x}$, which is a function of $\boldsymbol{\alpha}$, back into the Lagrangian. Thus we have for the dual function

$$g(\boldsymbol{\alpha}) = L(\mathbf{x}(\boldsymbol{\alpha}), \boldsymbol{\alpha}) \tag{20}$$

with $\mathbf{x}(\boldsymbol{\alpha}) = \arg\min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\alpha})$. The dual function $g(\boldsymbol{\alpha})$ is concave. For the convex optimization problems with linear constraints considered here, the optimal value $\tilde{\boldsymbol{\alpha}}$ for $\boldsymbol{\alpha}$ can be found by maximizing the dual function subject to $\boldsymbol{\alpha}_{ineq} \geq \mathbf{0}$ (see, e.g. Boyd and Vandenberghe, 2004, ch. 5). If there are no inequality constraints this will be an unconstrained maximization which can be performed by solving $g'(\boldsymbol{\alpha}) = L'(\boldsymbol{x}(\boldsymbol{\alpha}), \boldsymbol{\alpha}) = \mathbf{0}$. For more general cases we will need to solve the *constrained dual problem*

$$\begin{array}{ll} \text{maximize} & g(\boldsymbol{\alpha}) \\ \text{subject to} & \boldsymbol{\alpha}_{ineq} \geq 0 \end{array} \tag{21}$$

In either case, when an optimal value $\tilde{\boldsymbol{\alpha}}$ has been found, the optimum value for $\mathbf{x}$ can be obtained from $\tilde{\mathbf{x}} = \mathbf{x}(\tilde{\boldsymbol{\alpha}})$.

The SPA uses a coordinate ascent method to maximize the dual function. This means that an iterative method is used that increases the dual function by successively changing one of the components of the dual vector at a time. If all components are updated one iteration is completed and a new one is started. After each change to the dual vector, the $\mathbf{x}$-vector will be updated. Let α_k^t denote the value of the k-th dual variable at iteration t, *after* it has been updated. Furthermore, let the entire dual vector at iteration t after this k-th updating be denoted by

$$\boldsymbol{\alpha}^{t,k} = (\alpha_1^t, ..., \alpha_k^t, \alpha_{k+1}^{t-1}, ..., \alpha_K^{t-1})^T.$$

Then the k-th component and the whole $\mathbf{x}$-vector are updated according to the following general scheme

$$\begin{array}{lll} \alpha_k - \text{update:} & \alpha_k^t = \arg\max_{\alpha_k} g(\alpha_1^t, ..., \alpha_{k-1}^t, \alpha_k, \alpha_{k+1}^{t-1}, ..., \alpha_K^{t-1}) \\ & \text{if} \quad k \in I_{ineq} \quad \text{then} \quad \alpha_k \geq 0 \\ \mathbf{x} - \text{update:} & \mathbf{x}^{t,k} = \arg\min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\alpha}^{t,k}) = \mathbf{x}(\boldsymbol{\alpha}^{t,k}) \end{array} \tag{22}$$

The algorithm is initialized by setting $\mathbf{x}$ equal to $\mathbf{x}_0$ which corresponds, for the adjustment models considered here, to initializing $\boldsymbol{\alpha}$ as $\boldsymbol{\alpha}^{0,0} = \mathbf{0}$ (compare (7), (8), (11)).

The iterations (22) can be described as follows. First the k-th coordinate of the current dual vector is updated by maximizing the dual function over this coordinate to arrive at a new dual vector $\boldsymbol{\alpha}^{t,k}$ wherein the k-th component has been changed so as to increase the dual function (hence, coordinate ascent). Then, if the k-th constraint is an inequality constraint, we truncate α_k^t to zero if a negative value would otherwise result. Finally, we update the whole $\mathbf{x}$-vector using the function $\mathbf{x}(\boldsymbol{\alpha})$ and the updated dual vector.

To implement the maximization step, we consider the gradient of $g(\boldsymbol{\alpha})$. For general D this gradient can be expressed as (using the chain rule)

$$\begin{aligned} g'(\boldsymbol{\alpha}) &= D'_{\mathbf{x}(\boldsymbol{\alpha})}(\mathbf{x}(\boldsymbol{\alpha}), \mathbf{x}_0)\partial\mathbf{x}(\boldsymbol{\alpha})/\partial\boldsymbol{\alpha} + (\mathbf{A}\mathbf{x}(\boldsymbol{\alpha}) - \mathbf{b}) + \mathbf{A}^T\boldsymbol{\alpha}\partial\mathbf{x}(\boldsymbol{\alpha})/\partial\boldsymbol{\alpha} \\ &= [D'_{\mathbf{x}(\boldsymbol{\alpha})}(\mathbf{x}(\boldsymbol{\alpha}), \mathbf{x}_0) + \mathbf{A}^T\boldsymbol{\alpha}]\partial\mathbf{x}(\boldsymbol{\alpha})/\partial\boldsymbol{\alpha} + \mathbf{A}\mathbf{x}(\boldsymbol{\alpha}) - \mathbf{b} \\ &= \mathbf{A}\mathbf{x}(\boldsymbol{\alpha}) - \mathbf{b}, \end{aligned} \tag{23}$$

where the last line follows from the previous one because the term within square brackets is the gradient $L'_{\mathbf{x}}(\mathbf{x}, \boldsymbol{\alpha})$ evaluated at $\mathbf{x}(\boldsymbol{\alpha})$ and hence zero by the definition of $\mathbf{x}(\boldsymbol{\alpha})$. Thus, an unconstrained maximum in the k-th coordinate of the dual function occurs when α_k is such that the residual of the corresponding constraint is zero, obtained by solving $g'_k = 0$ for α_k, which yields the value for α_k^t in (22). For inequality constraints the dual function is maximized over non-negative α_k and the residuals need not to be zero.

Using these results, we will now specialize the general algorithmic scheme (22) to algorithms for the (weighted) least squares and Kullback-Leibler discrepancy functions.

SPA for least squares. For the least squares criterion we have, from adjustment model (7), that the function $\mathbf{x}(\boldsymbol{\alpha})$ evaluated in $\boldsymbol{\alpha}^{t,k-1}$ is given by $\mathbf{x}(\boldsymbol{\alpha}^{t,k-1}) = \mathbf{x}_0 - \sum_{l=1}^{k-1} \mathbf{a}_l \alpha_l^t - \sum_{l=k}^{K} \mathbf{a}_l \alpha_l^{t-1}$. So, if we change α_k^{t-1} to α_k^t to obtain $\boldsymbol{\alpha}^{t,k}$, we have

$$\mathbf{x}(\boldsymbol{\alpha}^{t,k}) = \mathbf{x}(\boldsymbol{\alpha}^{t,k-1}) - \mathbf{a}_k(\alpha_k^t - \alpha_k^{t-1}), \tag{24}$$

which becomes the $\mathbf{x}$-update for the least squares case of the general algorithm (22). With this expression for $\mathbf{x}(\boldsymbol{\alpha}^{t,k})$, solving $g_k' = 0$ amounts to

$$\alpha_k^t = \alpha_k^{t-1} + (\mathbf{a}_k^T \mathbf{x}(\boldsymbol{\alpha}^{t,k-1}) - b_k)/\mathbf{a}_k^T \mathbf{a}_k, \tag{25}$$

which becomes the α_k-update for the least squares case.

For equality constraints this algorithm can be simplified because the constraint $\alpha_k \geq 0$ of (22) does not apply and we can substitute $\alpha_k^t - \alpha_k^{t-1}$ obtained from (25) into the $\mathbf{x}$-updating equation (24), resulting in the single-line algorithm

$$\mathbf{x}^{t,k} = \mathbf{x}(\boldsymbol{\alpha}^{t,k}) = \mathbf{x}^{t,k-1} - \mathbf{a}_k(\mathbf{a}_k^T \mathbf{x}^{t,k-1} - b_k)/\mathbf{a}_k^T \mathbf{a}_k, \tag{26}$$

showing that for equality constraints the dual variables need not be calculated to find the optimal adjusted $\mathbf{x}$-vector. They could still be calculated however, since, as parameters of the adjustment model, they may be of interest themselves. For inequality constraints the $\mathbf{x}$-update is also given by (26) for positive α_k^t and is given by (24) with $\alpha_k^t = 0$ otherwise.

Notice that a positive dual variable for an inequality constraint can only arise if at some time during the iterations this constraint is violated, leading to a positive residual, say, $\mathbf{a}_k^T \mathbf{x}^{t,k-1} - b_k > 0$. The updated value of the dual variable, α_k^t will then set the residual to zero so that the constraint is satisfied with equality, since $\mathbf{a}_k^T \mathbf{x}^{t,k} = \mathbf{a}_k^T \mathbf{x}^{t,k-1} - \mathbf{a}_k^T \mathbf{a}_k (\mathbf{a}_k^T \mathbf{x}^{t,k-1} - b_k)/\mathbf{a}_k^T \mathbf{a}_k = b_k$. Now, suppose that a positive α_k value occurs at iteration t, then the contribution to the adjustment by the corresponding constraint is given by $-\mathbf{a}_k \alpha_k^t$. Next, suppose that at iteration $t+1$ the constraint has become satisfied with slack, so that the residual is *strictly* smaller than zero, due to changes to $\mathbf{x}$ that have occurred in between. The dual variable will still be updated, by (25), and the corresponding updated contribution to the adjustment is given by $\mathbf{x}^{t+1,k} - \mathbf{x}_0 = -\mathbf{a_k} \alpha_k^{t+1}$, where $0 < \alpha_k^{t+1} < \alpha_k^t$ by (25). In other words, on account of the k-th constraint, $\mathbf{x}^{t+1,k}$ is moved back closer towards to $\mathbf{x}_0$ compared to $\mathbf{x}^{t,k}$. This shows that a simpler intuitive approach that only adjusts $\mathbf{x}$ when an inequality constraint is violated will not always lead to an optimal solution, because the possibility of improving the objective function of the minimization problem by 'removing' some of the adjustment made in previous iterations (with a slack in the current iteration) is not exploited.

Using the adjustment model (8) for $\mathbf{x}(\boldsymbol{\alpha})$, it is straightforward to show that the SPA for weighted least squares amounts to replacing $\mathbf{a}_k^T \mathbf{a}_k$ by $\mathbf{a}_k^T Diag(\mathbf{w})^{-1} \mathbf{a}_k$ in (25) and premultiplying $\mathbf{a}_k(\alpha_k^t - \alpha_k^{t-1})$ by $Diag(\mathbf{w})^{-1}$ in (24).

SPA for the KL criterion. For the KL-divergence we have, from adjustment model (11), that the function $\mathbf{x}(\boldsymbol{\alpha})$ evaluated at $\boldsymbol{\alpha}^{t,k-1}$ is given by $\mathbf{x}(\boldsymbol{\alpha}^{t,k-1}) = \mathbf{x}_0 \times \prod_{l=1}^{k-1} \exp(-\mathbf{a}_l \alpha_l^t) \times \prod_{l=k}^{K} \exp(-\mathbf{a}_l \alpha_l^{t-1})$. So, if we change α_k^{t-1} to α_k^t to obtain $\boldsymbol{\alpha}^{t,k}$, we have

$$\mathbf{x}(\boldsymbol{\alpha}^{t,k}) = \mathbf{x}(\boldsymbol{\alpha}^{t,k-1}) \times \exp\{\mathbf{a}_k(\alpha_k^{t-1} - \alpha_k^t)\}, \tag{27}$$

with $\times$ and $\prod$ denoting element-wise products and 'exp' denoting element-wize exponentiation. Expression (27) becomes the $\mathbf{x}$-update equation for the KL-case of the general algorithm (22). Contrary to the least squares case, there is now no explicit solution of $g_k' = 0$ for α_k in general (but see below for an important exception). However, we can increase the dual function by taking a uni-dimensional Newton step instead, i.e. by updating α_k according to

$$\begin{aligned} \alpha_k^t &= \alpha_k^{t-1} - g_k'(\boldsymbol{\alpha}^{t,k-1}) / g_k''(\boldsymbol{\alpha}^{t,k-1}) \\ &= \alpha_k^{t-1} + (\mathbf{a}_k^T \mathbf{x}(\boldsymbol{\alpha}^{t,k}) - b_k) / \mathbf{a}_k^T Diag(\mathbf{x}^{t,k-1}) \mathbf{a}_k, \end{aligned} \tag{28}$$

with $g_k'(\boldsymbol{\alpha}^{t,k-1})$ and $g_k''(\boldsymbol{\alpha}^{t,k-1})$ the first and second derivatives of the dual function with respect to the k-th component of $\boldsymbol{\alpha}^{t,k-1}$. For equality constraints we can, just as in the least squares case, simplify this algorithm by substituting $\alpha_k^t - \alpha_k^{t-1}$ obtained from (28) into the $\mathbf{x}$-updating equation (27).

An explicit solution of $g_k' = 0$ can be obtained if the elements of $\mathbf{a}_k$ are all either one or zero. This corresponds to the case where the constraint k prescribes that the sum of a number of adjustable variables is equal to a constant b_k, which may be an unadjustable variable or a function of unadjustable variables. We will denote the sum of the x-variables corresponding to constraint k by X_{+k}, that is $X_{+k} = \mathbf{a}_k^T \mathbf{x}$. Then, using (23) and (27) we can write

$$g_k'(\boldsymbol{\alpha}^{t,k}) = \mathbf{a}_k^T \mathbf{x}(\boldsymbol{\alpha}^{t,k}) - b_k = \mathbf{a}_k^T \mathbf{x}(\boldsymbol{\alpha}^{t,k-1}) \times \exp\{\mathbf{a}_k(\alpha_k^{t-1} - \alpha_k^t)\} - b_k = 0,$$

and hence, we obtain for the α_k-update for this specific kind of constraints

$$\exp(\alpha_k^t) = \exp(\alpha_k^{t-1}) \frac{b_k}{X_{+k}^{t,k-1}} \ . \tag{29}$$

As before, for equality constraints we can obtain a single-line algorithm by combining (29) and (27) resulting in the updating equation

$$\begin{aligned} x_j^{t,k} &= x_j^{t,k-1} \frac{b_k}{X_{+k}^{t,k-1}}, && \text{for } a_{kj} = 1 \\ &= x_j^{t,k-1}, && \text{for } a_{kj} = 0. \end{aligned}$$

This is the same proportional adjustment as used by the IPF-algorithm that, when applied to a rectangular contingency table, adjust the counts in the table to new row- and column-totals by multiplying, successively, the counts in each row by a factor such that they add up to the new row-total and similarly for the columns. Of course, the SPA for the Kullback-Leibler criterion is much more general than this special case because the SPA

- handles more general data structures rather than only square tables
- it is not limited to constraint matrices with elements all equal to either zero or one
- is designed to handle inequality constraints as well as equations.