COMPUTATIONAL STATISTICS & DATA ANALYSIS

ELSEVIER Computational Statistics & Data Analysis 30 (1999) 359–379

A least squares algorithm for a mixture model for compositional data

Ab Mooijaart[a, *], Peter G.M. van der Heijden[b], L. Andries van der Ark[b]

[a] *Department of Psychology, Faculty of Social Sciences, University of Leiden, Wassenaarseweg 52, 2333 AK, Leiden, The Netherlands*

[b] *Department of Methodology and Statistics, Faculty of Social Sciences, University of Utrecht, Utrecht, The Netherlands*

Received 1 December 1997; received in revised form 1 August 1998; accepted 30 November 1998

Abstract

The estimation of a model for compositional data is studied where the data are approximated by a mixture of latent compositions. This model is variously known as “endmember analysis” or “latent budget analysis”. Two estimation procedures are available. The first uses a procedure which is incorrect in the sense that, although it claims to be a least squares procedure, it does not always minimize a least squares criterion. The second uses a maximum likelihood procedure starting from assumptions that are often violated for compositional data. In this paper we propose a constrained (weighted) least squares procedure for the estimation of the model.

Keywords: Latent budget model; End member model; Mixture model; Compositional data; EM algorithm; Weighted least squares algorithm

1. Introduction

There are two main approaches to the analysis of compositional data. One approach is worked out in considerable detail by Aitchison (1986), who transforms compositional data into a log-ratio covariance matrix. He then proposes models in terms of this covariance matrix, and shows how all sorts of questions about compositional data can be answered in this way. In a second approach, which we will study in this paper, the compositional data are approximated by a mixture of K latent, i.e.

* Corresponding author.

0167-9473/99/$ – see front matter
PII: S 0167-9473(98)00098-X

unknown, compositions, that have to be estimated from the data. In geology this model is called “endmember analysis”, where the unknown compositions are called “endmembers” (see, e.g. Renner, 1993a,b; Weltje, 1997). In the social sciences this model is known as “latent budget analysis”, where the unknown compositions are called “latent budgets” (see, e.g. Leeuw and van der Heijden, 1988).

A typical example of data from the social sciences that lead to compositional data is given in Table 1. Table 1 shows the time allocation in the Netherlands for a week, in minutes, cross-classified by gender (2 levels), age (5 levels) and year (3 levels). Thus there are $2 \times 5 \times 3 = 30$ rows. There are 18 main activities, such as paid work (1), sleeping (7), tv-radio (15) and the like. The data in Table 1 are obtained from three surveys (taken in 1975, 1980 and 1985) where respondents were asked to keep diaries (for more details, see Knulst and van Beek, 1990).

A row in Table 1 becomes a composition if the elements in this row are divided by the row total (in this case 10 080 min). Thus the row elements become proportions adding up to one. This leads to a matrix having 30 observed compositions.

The model to be discussed in this paper estimates K latent compositions, where K is to be determined either from theory or from the data. Each of the 30 observed compositions is approximated by a mixture of these K latent compositions. Since each of the observed compositions is approximated by a mixture (i.e. the approximation is a linear combination of the latent compositions with non-negative mixing proportions as weights adding up to 1), the latent compositions have to be “extreme” compositions. Thus for Table 1 each latent composition could be interpreted as an “extreme” way of allocating the time. The model approximates each of the 30 observed compositions by a mixing of these extreme ways of allocating time. In fact, in their original paper de Leeuw and van der Heijden (1988) proposed this model for the analysis of time budgets, and they called it “the latent time budget model”.

A typical problem for endmember analysis is to unravel the mixed provenance of coastal sands (see Weltje, 1997). The idea is that the coastal sands are mixtures of sediments derived from different sources, where the exact nature of the mixing process is unknown. Thus the model assumes that the observed compositions derived from sediments are approximated by mixtures of sediments of K sources.

The two traditions that employ this mixture model, endmember analysis and latent budget analysis, have different ways of approximating the observed compositions by expected compositions. For endmember analysis, Renner (1993a,b) proposed a least squares procedure to approximate the observed compositions by expected compositions. In the first step of this procedure a singular value decomposition of the matrix of observed compositions is carried out, yielding a lower rank approximation of this matrix, where this approximation is optimal in a least squares sense. However, since this lower rank matrix can have negative elements, and the rows of the approximation do not add up to one, Renner subsequently made adjustments to this matrix to ensure that the elements of this matrix are non-negative and each row sums to one. The resulting matrix is considered to be the matrix with estimates of expected compositions. Thus, as has been pointed out by van der Heijden (1994), this final matrix no longer approximates the observed matrix in a least squares sense.

For latent budget analysis, it has been assumed thus far that the observed matrix with compositions is derived from a matrix with frequencies sampled from a product-multinomial distribution (see van der Heijden et al., 1992). The model is then estimated by maximum likelihood using an EM-algorithm. de Leeuw and van der Heijden (1988) discussed an example where this assumption is realistic: the proportions in time budgets were derived using the so-called random spot check method. This method is quite popular in anthropology. The basic idea is that independent observations are made of a person's behaviour by checking what he is doing at random points in time; when it turns out that he is eating during 5 out of the 25 observations, he is assumed to eat 20% of his time. If the observations are indeed independent, then the assumption of a product-multinomial distribution is realistic. However, for many other applications the assumption of a product-multinomial distribution is not at all realistic. To continue with the same topic of time budgets (though many other examples could be given), the minutes in Table 1 are derived from diaries, and it is immediately clear that the minutes cannot be considered realisations from a product-multinomial distribution.

In conclusion, there are compositional data for which no acceptable estimation procedure for the model under study is available. For example, when the time budgets are derived from diaries, the least squares procedure proposed by Renner (1993a,b) does not lead to a proper least squares approximation of the matrix observed compositions, and it is evident that the minutes do not follow a product-multinomial distribution, so that assumptions for the maximum-likelihood procedure of van der Heijden et al. (1992) are not fulfilled.

This shows that there are at least two ways for the further development of estimation procedures for mixture model under study. One way is to extend the maximum-likelihood procedure to different sampling models; the other way is to deal with the problems in the current least squares procedure. In this paper we concentrate on the development of a sound least squares estimation procedure of the latent budget model. We have chosen this last option because we can analyze any compositional data set with a sound least squares algorithm, whereas extension of the maximum likelihood procedure to a larger set of sampling models leaves open the possibility that for a particular data set the assumptions of all sampling models may be violated. Choice for a least squares procedure can be motivated by the fact that in such a procedure the matrix of observed data is approximated by a lower rank matrix in a weighted least squares sense.

In Section 2 we begin with a formal presentation of the model, and we define a general least squares function for the model. In Section 3 we outline the way constrained least squares problems can be solved in general, and this procedure is then applied to our problem in Section 2. Section 4 discusses the estimation of the model parameters by minimizing a weighted least squares function. Section 5 discusses the example mentioned in the Introduction. In Section 6 several algorithmic issues, such as convergence properties, choice of starting values and computational efficiency are discussed and illustrated by analyzing several data sets. An important issue in this section is the discussion of the existence of local optima. It is well known from optimizing the likelihood function in the latent budget model that local

Table 1
Time allocation in the Netherlands, cross-classified by Gender, Age and Year (rows of the table). There are 18 activities (columns of the table). In a cell the average number of minutes in a day is given that people of a certain gender and age spent in either 1975, 1980 or 1985 to some activity. There are 18 activities, namely, (1) paid work, (2) domestic work, (3) caring for members household, (4) shopping, (5) personal need, (6) eating and drinking, (7) sleeping and resting, (8) education, (9) participation in volunteer work, (10) social contacts, (11) going out, (12) sports, hobbies, games, (13) gardening, taking care of pets, (14) recreation outside, (15) tv, radio, audio (16) reading, (17) relaxing, and (18) others

			Activities																	
Gender	Age	Year	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
M	12–24	75	901	87	33	120	289	508	3737	1447	128	515	490	419	111	48	752	272	78	146
		80	769	157	28	138	294	528	3765	1455	101	505	396	436	102	41	815	256	56	240
		85	707	155	15	127	316	527	3744	1537	92	449	441	485	100	64	860	188	73	200
	25–34	75	2180	250	194	152	293	623	3380	124	129	609	382	269	173	69	700	366	64	124
		80	1992	269	206	157	316	649	3403	245	126	649	321	279	213	35	671	318	58	172
		85	1899	341	184	183	302	605	3397	208	143	599	391	271	231	67	812	243	57	148
	35–49	75	1901	249	99	173	351	660	3463	56	195	671	360	206	259	88	785	316	59	188
		80	2008	289	128	157	339	709	3445	90	156	593	240	280	238	45	804	343	44	170
		85	2093	331	136	185	332	650	3347	85	148	479	336	291	268	64	812	319	58	146
	50–64	75	1708	244	51	227	350	709	3560	18	122	603	237	209	256	116	921	468	79	203
		80	1357	337	54	221	364	744	3569	58	207	704	279	299	288	76	862	413	57	190
		85	1206	450	25	230	352	686	3533	46	272	554	264	316	309	112	1012	467	68	174
	> 65	75	176	617	124	273	365	763	3801	10	159	811	213	297	366	86	1161	477	157	223
		80	71	563	27	251	392	767	3871	43	192	671	220	403	312	117	1198	660	92	230
		85	95	636	38	264	383	707	3694	54	214	619	274	476	308	178	1233	578	104	225

Table 1 (Continued).

			Activities																	
Gender	Age	Year	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
F	12–24	75	723	494	135	208	359	536	3744	1163	125	592	364	348	90	32	594	292	73	208
		80	665	460	99	200	377	513	3777	1321	88	557	400	370	76	32	581	257	74	234
		85	564	397	86	223	387	495	3821	1436	80	527	396	352	86	41	702	207	63	214
	25–34	75	439	1342	635	347	311	593	3526	77	85	780	316	306	149	41	547	300	88	199
		80	471	1338	673	336	339	607	3532	115	115	776	270	352	131	32	497	275	54	167
		85	704	1147	651	336	337	572	3447	120	115	736	303	368	145	44	565	265	63	164
	35–49	75	299	1567	296	372	325	664	3567	104	133	694	225	335	198	37	622	356	76	207
		80	375	1605	309	347	346	633	3554	98	143	689	229	440	154	34	576	311	63	174
		85	412	1529	308	373	351	656	3444	68	196	699	277	453	170	45	582	307	59	153
	50–64	75	151	1600	83	376	367	601	3673	27	195	758	255	323	197	53	710	478	78	154
		80	153	1558	84	335	368	613	3701	30	179	810	212	504	190	41	644	390	53	212
		85	233	1487	82	352	385	595	3566	40	195	721	268	545	217	76	708	377	64	170
	> 65	75	11	1319	78	384	372	635	3849	6	108	929	219	297	169	37	888	485	63	230
		80	6	1409	154	292	453	665	3713	21	124	796	187	482	191	39	860	404	67	216
		85	19	1318	44	320	366	615	3675	23	139	749	202	579	169	52	1076	460	69	204

optima may exist. These local optima have different function values, so they are really local optima do not arise from the fact that the model may be not identified. We end with some concluding remarks in Section 7.

2. Presentation of the model and the estimation problem

Let $\boldsymbol{P}$ be an $I \times J$ matrix with observed compositional data. This matrix consists of I compositions, each composed from the same J elements, but in different proportions. The elements of composition i $(i=1,\dots,I)$, denoted as p_{ij} $(i=1,\dots,I,\ j=1,\dots,J)$, are non-negative and add up to one: $\sum_j p_{ij} = 1$. The model examined in this paper approximates the rows of the matrix $\boldsymbol{P}$ by a mixture of K unknown compositions, collected in a $J \times K$ matrix $\boldsymbol{B}$, with elements b_{jk} nonnegative and adding up to one for each unknown composition k: $\sum_j b_{jk} = 1$. The mixing parameters for composition i are collected in a matrix $\boldsymbol{A}$, with elements a_{ik} nonnegative and adding up to one for each composition i: $\sum_k a_{ik} = 1$. Thus the model is

$$\Pi = \boldsymbol{AB}', \tag{1}$$

where the matrix Π approximates the matrix $\boldsymbol{P}$. de Leeuw and van der Heijden (1991) have shown that if $K = \min(I,J)$, $\boldsymbol{P} = \Pi$. If $K < \min(I,J)$, Π provides a rank K approximation of $\boldsymbol{P}$.

Model (1) is unidentified, since $\Pi = \boldsymbol{AB}' = (\boldsymbol{AT})(\boldsymbol{T}^{-1}\boldsymbol{B}') = \boldsymbol{A}^*\boldsymbol{B}^{*'}$. Once estimates for $\boldsymbol{A}$ and $\boldsymbol{B}$ are found, a proper choice of $\boldsymbol{T}$ has to be made, in order to solve the identification problem. This has been worked out by van der Ark et al. (in press) in considerable detail. They have two proposals for $\boldsymbol{T}$ to choose from. In what they call *the outer extreme solution* $\boldsymbol{T}$ is chosen such that the latent budgets in $\boldsymbol{B}^*$ become as different as possible; this makes the elements in each row of $\boldsymbol{A}^*$ more similar. In what they call *the inner extreme solution* $\boldsymbol{T}$ is chosen such that the elements in each row of $\boldsymbol{A}^*$ are as different as possible; this makes the latent budgets in $\boldsymbol{B}^*$ more similar. Thus $\boldsymbol{T}$ is used in a similar way as in factor analysis, where matrices of factor loadings are rotated to simple structure. We refer to van der Ark et al. (in press) for more details.

Relations of this model with latent class analysis and correspondence analysis are summarized in van der Heijden et al. (1992) and van der Ark and van der Heijden (1998). In fact, Goodman (1974) introduced the latent budget model in sociology as a reparametrization of latent class analysis of a two-way contingency table. For the properties of the model, extensions, and possible applications, we refer to de Leeuw et al. (1991), van der Heijden et al. (1992), Renner (1993a,b), Siciliano and van der Heijden (1994), Weltje (1997), van der Ark and van der Heijden (1998) and van der Ark et al. (in press), and the papers cited there. Ideas similar to those in the latent budget model are also found in archetypal analysis (Cutler and Breiman, 1994).

To fit the model, we propose to minimize the least squares loss function developed below. To summarize, the model is $\Pi = \boldsymbol{AB}'$, where the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ have orders $(I \times K)$ and $(J \times K)$, respectively, where each row of $\boldsymbol{A}$ and each column of $\boldsymbol{B}$ sum to 1. All elements of $\boldsymbol{A}$ and $\boldsymbol{B}$ must be non-negative. An observed matrix $\boldsymbol{P}$ is to be approximated by a matrix Π. Therefore a loss function will be defined which must be minimized with respect to the unknown parameters, i.e. the elements of the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. For this we choose a weighted least squares function.

In this weighted least squares function we allow for weights for the rows and for weights for the columns. For the rows we introduce a diagonal weight matrix $\boldsymbol{V}$ having elements v_i, and for the columns we introduce a diagonal matrix $\boldsymbol{W}$ having elements w_j. The function to be minimized is then

$$f_{\boldsymbol{AB}} = \mathrm{SSQ}[\boldsymbol{V}(\boldsymbol{P} - \boldsymbol{AB}')\boldsymbol{W}], \tag{2}$$

where SSQ[$\boldsymbol{X}$] is defined as the sum of squares of the elements of matrix $\boldsymbol{X}$. The side conditions which hold are

$$\boldsymbol{A}\mathbf{1}_K = \mathbf{1}_I, \tag{3a}$$

$$\mathbf{1}_{J'}\boldsymbol{B} = \mathbf{1}_{K'}, \tag{3b}$$

$$a_{ik}, b_{jk} \geq 0 \quad (i = 1,\ldots,I;\ j = 1,\ldots,J;\ k = 1,\ldots,K), \tag{3c}$$

where $\mathbf{1}_N$ is defined as a $(N \times 1)$ column vector with unit elements only.

In the context of compositional data an appropriate choice for elements v_i could depend on the extent of data n_i from which composition i is derived. Consider the example in Table 1. If one row were based on twice as much respondents as another row, then we might be willing to give the former row two times as much weight in the determination of the final solution. In this case one should choose $v_i = p_i^{1/2}$. Thus, a difference between $\boldsymbol{P}$ and $\boldsymbol{AB}'$ in the former row would count twice as much in $f_{\boldsymbol{AB}}$ as a comparable difference in the latter row.

An appropriate choice for elements w_j could depend on the relative sizes p_j of the column margins. For instance, assume that we want to compare the elements of two columns of the data matrix, one with an average proportion of 0.20 and one with an average proportion of 0.02. Then, in principle, the observed elements for the column with an average proportion of 0.20 can differ much more than the one with an average proportion of 0.02, and an unweighted solution would be based mainly on the columns with larger p_j. Assume that we would like to weight a difference between $\boldsymbol{P}$ and $\boldsymbol{AB}'$ in the column with proportion 0.02 ten times as much as a comparable difference in the column with proportion 0.20. In this case one could choose, e.g., $w_j = 1/p_j^{1/2}$ or the inverse variance $w_j = 1/(p_j(1 - p_j))$.

We note that similar weights are used in the estimation phase of correspondence analysis when a singular value decomposition is made of the matrix of row profiles.

3. Minimizing the general least squares function

Instead of minimizing Eq. (2) under the conditions (3) directly, we propose an algorithm where $\boldsymbol{A}$ and $\boldsymbol{B}$ are estimated alternately. There are two reasons for doing this proposal: in the first place there are no conditional closed-form solutions, i.e. $\boldsymbol{A}$ given $\boldsymbol{B}$ and $\boldsymbol{B}$ given $\boldsymbol{A}$. Secondly, the advantage of our algorithm is that, instead of minimizing the complicated nonlinear function in Eq. (2) under the conditions (3), we minimize simple least squares functions, only. Therefore, in this section we discuss how a linear set of equations can be solved by a general least squares function. Then, in Section 4, we show how our minimization problem for solving the parameters of the latent budget model can be transformed into this general function. The general least squares function for a linear set of equations can be written as

$$f_x = \mathrm{SSQ}[\boldsymbol{Qx} - \boldsymbol{r}], \tag{4a}$$

with side conditions

$$\boldsymbol{C}_1\boldsymbol{x} = \boldsymbol{d}_1, \tag{4b}$$

$$x_i \geq 0, \tag{4c}$$

where all matrices and vectors are of appropriate orders. This is a well-known problem and a detailed discussion how to find solutions can be found in Lawson and Hanson (1995) and Björck (1996). If the side conditions were only $x_i \geq 0$, then the algorithm NNLS (Non-Negative Least Squares) of Lawson and Hanson or the active set algorithm for problem BLS (Least Squares with simple Bounds) of Björck could be used. However, in the latent budget model some sub-sets of parameters should add to one, so there are side conditions of the form $\boldsymbol{C}_1\boldsymbol{x} = \boldsymbol{d}_1$. Therefore we shall discuss how to expand the NNLS or BLS algorithm such that Eq. (4b) holds. Three methods will be discussed. In the first method a penalty function will be added to the least squares function. In the second method "an active constraint method" will be given. In the third method, a branch and bound algorithm will be discussed.

Method 1 (*Least squares function with a penalty function*)**.** Here the least squares function we minimize is

$$f_x = \mathrm{SSQ}\left[\begin{pmatrix} \boldsymbol{Q} \\ \cdots \\ m\boldsymbol{C}_1 \end{pmatrix}\boldsymbol{x} - \begin{pmatrix} \boldsymbol{r} \\ \cdots \\ m\boldsymbol{d}_1 \end{pmatrix}\right] \tag{5a}$$

with side condition

$$x_i \geq 0. \tag{5b}$$

In Eq. (5a) m is an arbitrary known scalar. To have more insight in this equation we can write f_x also as

$$f_x = \mathrm{SSQ}[\boldsymbol{Qx} - \boldsymbol{r}] + m^2\mathrm{SSQ}[\boldsymbol{C}_1\boldsymbol{x} - \boldsymbol{d}_1]. \tag{5c}$$

If m is chosen "large", then the solution of $\boldsymbol{x}$ by minimizing f_x under the conditions $x_i \geq 0$, will satisfy approximately also the restrictions $\boldsymbol{C}_1\boldsymbol{x} = \boldsymbol{d}_1$. Furthermore, the larger m will be chosen, the "better" this approximation is. The second term on the right-hand side of Eq. (5c) is called a penalty function. Because $\mathrm{SSQ}[\boldsymbol{C}_1\boldsymbol{x} - \boldsymbol{d}_1]$ has a minimum of zero it follows that if m is taken "large", e.g. $m = 10\,000$, then there is a heavy penalty if $\boldsymbol{C}_1\boldsymbol{x} \neq \boldsymbol{d}_1$. So solving $\boldsymbol{x}$ from Eqs. (5a) and (5b) will be "close" to the solution of $\boldsymbol{x}$ from Eq. (4). There are two remarks to be made here. First, the solution of $\boldsymbol{x}$ from minimizing Eq. (5a) will only be an approximation of the solution of $\boldsymbol{x}$ by minimizing Eq. (4). How close the two solutions are depend on the problem we are investigating. Second, there is an disadvantage of this method. In general it holds that if $\boldsymbol{Q}$ is of full column rank, then there is a unique solution of $\boldsymbol{x}$ from minimizing the least squares function as defined in Eq. (4). However, if m is large, then in general the matrix $(\boldsymbol{Q}';\ m\boldsymbol{C}_1')'$ will be poorly conditioned because the magnitude of the elements of this matrix may become very different. In such a case a solution of $\boldsymbol{x}$ may become very unstable. See for a discussion of this topic Lawson and Hanson (1995), paragraph 22. Furthermore, for solving the parameters, collected in the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, in the models we are discussing in this paper we have an iterative process in which alternatingly matrix $\boldsymbol{A}$ and $\boldsymbol{B}$ will be solved. So the least squares solution of $\boldsymbol{x}$ has to be solved many times. This means, in particular near the optimum, convergence of the parameters may become problematic because the parameters in $\boldsymbol{A}$ and $\boldsymbol{B}$ are estimated only approximately. Therefore we do not, in opposite to Cutler and Breiman (1994), recommend this method.

Method 2 (*Active constraint method*)**.** For a more detailed discussion of this method, we refer to Gill and Murray (1974) and Zangwill (1969). The idea of this method is as follows: each inequality constraint in Eq. (4c) defines a boundary. For the optimal solution of $\boldsymbol{x}$, either all elements are positive, or one or more elements are equal to zero. If $\boldsymbol{x}$ lies on several boundaries at the same time, i.e. there is a set of active constraints for which $x_i = 0$, then we can write

$$\boldsymbol{C}_2\boldsymbol{x} = \boldsymbol{0}. \tag{6}$$

Combining Eqs. (4b) and (4c) gives

$$\boldsymbol{C}\boldsymbol{x} = \boldsymbol{d}, \tag{6a}$$

where

$$\boldsymbol{C} = \begin{pmatrix} \boldsymbol{C}_1 \\ \cdots \\ \boldsymbol{C}_2 \end{pmatrix}, \tag{6b}$$

$$\boldsymbol{d} = \begin{pmatrix} \boldsymbol{d}_1 \\ \cdots \\ 0 \end{pmatrix}. \tag{6c}$$

Using the Lagrange function, we have to minimize

$$f_{x,\lambda} = \mathrm{SSQ}[\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{r}] - 2\lambda'(\boldsymbol{C}\boldsymbol{x} - \boldsymbol{d}), \tag{7}$$

where λ is a vector of Lagrange multipliers. Defining the unrestricted solution of $\boldsymbol{x}$ as

$$\boldsymbol{x}^{\mathrm{O}} = (\boldsymbol{Q}'\boldsymbol{Q})^{-1}\boldsymbol{Q}'\boldsymbol{r}. \tag{8}$$

Taking the derivatives of $f_{x,\lambda}$ with respect to $\boldsymbol{x}$ and λ and solving for λ gives

$$\lambda = (\boldsymbol{C}(\boldsymbol{Q}'\boldsymbol{Q})^{-1}\boldsymbol{C}')^{-1}(\boldsymbol{d} - \boldsymbol{C}\boldsymbol{x}^{\mathrm{O}}), \tag{9a}$$

which gives as a solution for $\boldsymbol{x}$

$$\boldsymbol{x} = \boldsymbol{x}^{\mathrm{O}} + (\boldsymbol{Q}'\boldsymbol{Q})^{-1}\boldsymbol{C}'\lambda. \tag{9b}$$

If this solution is not feasible, then new constraints are activated, i.e. new rows are added to matrix $\boldsymbol{C}_2$. A new row consists of all zeros except for one '1', indicating the element that has to be constrained to zero. This process is repeated until a feasible solution is achieved.

The reason we give the general formulation of the problem is that in the case with inequality restrictions the Lagrange multipliers have to be computed to check whether a temporary solution is a final optimal solution or not. An important property of the Lagrange multipliers is that if a constraint corresponds to an inequality constraint, then the corresponding Lagrange multiplier should be non-negative (see, e.g. pp. 159 –160 of Lawson and Hanson (1995)). This Lagrange multiplier can be computed by taking the derivatives of $f_{x,\lambda}$ with respect to $\boldsymbol{x}$ and equating these derivatives to zero, this gives

$$\boldsymbol{Q}'(\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{r}) = \boldsymbol{C}'\lambda. \tag{10}$$

So λ can be computed from Eq. (10). Therefore, if we find a feasible solution and the smallest λ, corresponding to an inequality constraint, is negative, then the corresponding active constraint is made passive. This means that one row of $\boldsymbol{C}_2$ is dropped, and a new $\boldsymbol{x}$ will be solved from Eqs. (9a) and (9b). The algorithm thus consists of solving $\boldsymbol{x}$ from Eqs. (9a) and (9b) with different matrices $\boldsymbol{C}_2$, until $\boldsymbol{x}$ is feasible and all λ's, except the ones corresponding to the equality constraints as defined by $\boldsymbol{C}_1$, are non-negative. Note that the λ's corresponding to the equality restrictions as defined by $\boldsymbol{C}_1$ may be negative. This algorithm converges monotonically to a minimum, i.e. the function value decrease or remain equal and will never increase. See for a proof pp. 163–164 of Lawson and Hanson (1995). Although, this algorithm will converge, it may be to a local minimum (for more details, see Section 6).

Although Lawson and Hanson (1995) and Björck (1996) do give an algorithm for non-negative elements, they do not give explicitly an algorithm for the case with side condition $\boldsymbol{C}_1\boldsymbol{x} = \boldsymbol{d}_1$. Therefore we will give this algorithm in Appendix A.

Method 3 (*Branch and bound algorithm*)**.** Cutler (1993) discusses a branch and bound algorithm. In this algorithm a non-negative least squares algorithm is extended to the case of equality constraints. In the branch and bound algorithm a binary tree is constructed, with nodes corresponding to least squares sub-problems. The idea behind the branch and bound algorithm is that an optimal solution of the non-negative least squares problem can always be found by computing all possible sub-problems in which one or more elements are equal to zero. For instance, if the unconstrained solution is not feasible, then we can compute the solutions in which only one element is equal to zero. If all these solutions are not feasible, then we can compute all solutions in which all pairs or triples of elements are zero. Continuing this process will always lead to an optimal solution. A disadvantage of this method is that if the number of elements of $\boldsymbol{x}$ is large, then a huge number of problems have to be solved. For instance, if the number of elements in $\boldsymbol{x}$ is equal to m, then in the worst case we have to solve 2^{m-1} sub-problems. This is not very attractive. The branch and bound algorithm is some special searching strategy in order not to solve all the sub-problems. The key idea here is, again, that if a solution is feasible, and the Lagrange multipliers are non-negative, then further search in the direction of more restricted models is not needed anymore, because the Kuhn–Tucker conditions are fulfilled. Cutler (1993) discusses this method and compares this algorithm with the NNLS problem of Lawson and Hanson (1995). She finds that the branch and bound algorithm almost always outperforms the NNLS method. However, we think this conclusion is not fair, because she always starts the NNLS algorithm with the vector with zero elements only. This is indeed the starting vector of Lawson and Hanson, however, another starting vector may be much more optimal. For instance, Björck (1996) in his BLS algorithm, which is similar to the NNLS algorithm, starts with a starting vector which is feasible and the closest to the unconstrained solution. It is clear that this algorithm will converge in a smaller number of steps if the solution vector do not have to much zero elements.

Conclusion. We discussed three methods in this section. In the method with a penalty function the solution is only an approximation of the "true" solution and the solution may be unstable because of poorly condition of one of the matrices. Therefore we do not recommend this solution. The branch and bound method seems to work well in cases in which the number of elements of $\boldsymbol{x}$ is small. In cases in which this number is large, say 100 as may be the case in our latent budget model, it is better to start with a feasible vector which is closest to the unconstrained solution instead of starting with the zero vector. This in opposite the NNLS algorithm of Lawson and Hanson (1995) and Cutler (1993). However, in the case of a large number of elements in $\boldsymbol{x}$, it seems that the active constraint method needs less iteration steps than the branch and bound algorithm. Therefore, in this paper we will use the active constraint algorithm.

4. Estimation of the model parameters

In Section 3 we discussed how the general least squares function can be minimized. We now show how the least squares function (2) with side conditions (3) can be dealt with.

The function to be minimized can now be written as

$$f_{\boldsymbol{AB}} = \mathrm{SSQ}[\boldsymbol{V}(\boldsymbol{P} - \boldsymbol{AB}')\boldsymbol{W}], \tag{11}$$

where $\boldsymbol{V}$ and $\boldsymbol{W}$ are diagonal matrices. With matrix $\boldsymbol{V}$ the rows of the residual matrix, i.e. $\boldsymbol{P} - \boldsymbol{AB}'$, are weighted differently, whereas with $\boldsymbol{W}$ the columns are weighted differently. These matrices are known matrices and have to be specified by the researcher. Obviously, when $\boldsymbol{V}$ and $\boldsymbol{W}$ are specified as identity matrices, the function is equal to the unweighted least squares function. In other cases we speak of a weighted least squares function.

In the algorithm we propose, we estimate $\boldsymbol{A}$ and $\boldsymbol{B}$ alternately. This means that we first fix all the elements of $\boldsymbol{B}$ and minimize the least function by solving for the elements of $\boldsymbol{A}$, then we fix all elements of $\boldsymbol{A}$ and minimize the least function by solving for the elements of $\boldsymbol{B}$. This defines one cycle of the algorithm. By solving for $\boldsymbol{A}$ and $\boldsymbol{B}$ in each cycle, in this way, the function value will decrease or remain equal. Obviously, the function value will never increase. We repeat the number of cycles until a suitable convergence criterion is met (see Section 6). Note that in this way we apply an ordinary least squares (OLS) method at each step to solve for either $\boldsymbol{A}$ or $\boldsymbol{B}$ and by doing this alternately, we have defined a so-called alternating least squares (ALS) method. So, in the first step of each cycle we minimize, with respect to $\boldsymbol{A}$,

$$\begin{aligned} f_A &= \mathrm{SSQ}\,[\boldsymbol{V}(\boldsymbol{P} - \boldsymbol{AB}')\boldsymbol{W}] = \mathrm{SSQ}[\boldsymbol{W}(\boldsymbol{P}' - \boldsymbol{BA}')\boldsymbol{V}] \\ &= \mathrm{SSQ}[\boldsymbol{P}^{*\prime} - \boldsymbol{WBA}'\boldsymbol{V}] = \sum_{i=1}^{I} \mathrm{SSQ}\,[\boldsymbol{p}_{i.}^{*} - v_i \boldsymbol{WBa}_{i.}], \end{aligned} \tag{12}$$

where $\boldsymbol{P}^{*'}$ is equal to $\boldsymbol{WP}'\boldsymbol{V}$, $\boldsymbol{p}_{i.}^{*}$ is a column with elements the elements of row i of matrix $\boldsymbol{P}^{*}$, and $\boldsymbol{a}_{i.}$ is a column with elements the elements of row i of matrix $\boldsymbol{A}$. This function is in the appropriate form given in Eqs. (4a)–(4c) if we define $\boldsymbol{x} \equiv \boldsymbol{a}_{i.}$, $\boldsymbol{Q} \equiv v_i \boldsymbol{WB}$ and $\boldsymbol{r} \equiv \boldsymbol{p}_{i.}^{*}$. For the side conditions we have $\boldsymbol{C}_1 \equiv \mathbf{1}_{K'}$ and $\boldsymbol{d}_1 \equiv 1$, where vector $\mathbf{1}_K$ is a $(k \times 1)$ vector consisting of unit elements, only. Here $\boldsymbol{C}_1$ and $\boldsymbol{d}_1$ are used to specify that the elements of row i of matrix $\boldsymbol{A}$ add up to one.

In the second step of each cycle we minimize, with respect to $\boldsymbol{B}$,

$$\begin{aligned} f_B &= \mathrm{SSQ}\,[\boldsymbol{V}(\boldsymbol{P} - \boldsymbol{AB}')\boldsymbol{W}] = \mathrm{SSQ}\,[\boldsymbol{W}(\boldsymbol{P}' - \boldsymbol{BA}')\boldsymbol{V}] \\ &= \mathrm{SSQ}\,[\mathrm{vec}(\boldsymbol{P}^{*'}) - (\boldsymbol{VA} \otimes \boldsymbol{W})\mathrm{vec}(\boldsymbol{B})], \end{aligned} \tag{13}$$

where the "vec" operator in vec($\boldsymbol{B}$) is defined as an operator which stacks the elements of $\boldsymbol{B}$ columnwise in a column vector. This function is in the appropriate form given in (4a)–(4c) if we define $\boldsymbol{x} \equiv \mathrm{vec}(\boldsymbol{B})$, $\boldsymbol{Q} \equiv \boldsymbol{VA} \otimes \boldsymbol{W}$, and $\boldsymbol{r} \equiv \mathrm{vec}(\boldsymbol{P}^{*'})$. For the side conditions we have $\boldsymbol{C}_1 \equiv (\boldsymbol{I}_K \otimes \mathbf{1}_{J'})$ and $\boldsymbol{d}_1 \equiv \mathbf{1}_K$. Here $\boldsymbol{C}_1$ and $\boldsymbol{d}_1$ define that the elements of the K columns of matrix $\boldsymbol{B}$ add up to 1.

Appendix B discusses the solution of the model parameters without the use of Kronecker products, i.e. by using products of matrices of smaller order only.

5. An example

In this section we discuss the analysis of the data in Table 1. Since it would make no sense to assume a multinominal distribution and to apply a maximum likelihood estimation procedure, our weighted least squares approach seems reasonable. The weights used are $1/p_j^{1/2}$ for the columns (see Section 2.)

We carried out an analysis for $K = 1$, $K = 2$, $K = 3$, $K = 4$ and $K = 5$ latent budgets. For $K = 1$ $f_{AB} = 0.2306$. This is the baseline model, where all the estimates of expected compositions are equal. For $K = 2$ $f_{AB} = 0.1194$, and therefore (0.2306–0.1194)/0.2306 = 0.482 of the lack of fit for the baseline model is now modelled by taking two latent compositions instead of only one. For $K = 3$ $f_{AB} = 0.0399$, and now 0.827 of the lack of fit for the baseline model is modelled. For $K = 4$ $f_{AB} = 0.0136$ (proportion modelled is 0.941). When making a decision over the number of latent compositions we use a few guidelines. First, the proportion of lack of fit with respect to the baseline model should be as large as possible, so that the model gives an adequate description of the data. Second, the increase in fit due to inclusion of an extra latent composition should be large enough to justify the increased effort of interpreting the extra set of parameters estimated. Since the loss for $K=1$ equals 0.2306 and the loss for $K=18$ is zero, we could argue that by including an extra dimension we would like to reduce the loss by at least $0.2306/18=0.0128$ (this criterion is similar to the eigenvalue greater than one criterion in principal component analysis). And third, the estimates of the latent compositions and the estimated mixing parameters should be interpretable. Using these guidelines, we hesitated between three and four latent compositions. Mainly for reasons of space we will discuss only the solutions with two and three latent compositions (see Table 2).

The solutions are identified by the outer extreme solution, i.e. the latent budgets are made as dissimilar as possible (compare Section 2).

We begin with an interpretation of the parameter estimates of the model with two latent compositions. This should be seen only as an interpretation exercise, because the fit of the model with two latent composition is not very good. This model assumes that, in terms of time allocation, there are two extreme weeks, and all 30 observed (average) weeks in Table 1 are approximated by a mixture of these extremes. The estimates $\boldsymbol{B}$ for the extremes are most easily interpreted by comparing them with the average proportions, which are also the estimates for the model with $K=1$. This shows that the first latent composition (the first extreme) is characterized in particular by much more domestic work (0.1260 of the time versus 0.0779 on average) at the expense of paid work (0.0351 versus on average 0.0803). Smaller differences between the first latent composition and the average composition can be found for caring for household members, shopping, social contacts, which all happen more than average, and education, which happens less than average. The second composition is necessarily characterized in the opposite way, because the budget sizes times the latent budget elements should lead to the observed marginal proportions e.g. for paid work $0.6228\times0.0351+0.3772\times0.1540=0.0803$ (see de Leeuw et al. (1990) for a proof).

The 30 compositions derived from Table 1 are all approximated by mixture of both extremes. For example, in 1985 the average male time budget for a week is approximated for 0.35 from the first latent composition and for 0.65 from the second latent composition. For all ages, the first composition (characterized particularly by a high contribution of domestic work) is used more by females (especially from the age of 25) than by males. For males the use of this composition increases as they get older, particularly when they get 65 or older (the obligatory retiring age in the Netherlands). Interestingly, over the years the contribution of the first latent composition increases for males starting at the age of 25−64 (e.g. for males aged 50−64 from 45% to 59%), indicating that over the years there is an increasing participation in domestic work and a decreasing participation in paid work; for women aged 12–24 and 25–34 this contribution goes down.

The model with three latent compositions gives a much better account of the observed data, and we turn to its interpretation now. The first latent composition is now characterized by domestic work (0.1456) and an absence of paid work and education (both 0.0000). To a lesser extent there is also a higher contribution of caring, shopping and social contacts. The second latent composition is characterized by paid work (0.2363) and the absence of domestic work and education (both 0.0000). The third latent composition is characterized by education (0.2094) at the expense of paid and domestic work; there is also less time spent on caring, shopping, eating, social contacts, gardening and reading, and more time on sleeping, going out and sports. On average, the first latent component (characterized by domestic work) is used predominantly by women older than 24 and retired men, the second latent component (characterized by paid work) is used predominantly by men aged between 25 and 64,

Table 2
Weighted least squares estimates for the model parameters for the data in Table 1. Solutions are given for $K = 1$, $K = 2$ and $K = 3$ latent compositions

Model			$K = 1$	$K = 2$		$K = 3$		
mix.parameters			$k = 1$	$k = 1$	$k = 2$	$k = 1$	$k = 2$	$k = 3$
M	12–24	75	1.0000	0.0000	1.0000	0.0000	0.3115	0.6885
		80	1.0000	0.0052	0.9948	0.0547	0.2531	0.6922
		85	1.0000	0.0000	1.0000	0.0395	0.2254	0.7351
	25–34	75	1.0000	0.3110	0.6890	0.0826	0.8662	0.0512
		80	1.0000	0.3135	0.6865	0.1114	0.7853	0.1033
		85	1.0000	0.3500	0.6500	0.1493	0.7582	0.0925
	35–49	75	1.0000	0.3888	0.6112	0.1649	0.8030	0.0321
		80	1.0000	0.3724	0.6276	0.1473	0.8164	0.0363
		85	1.0000	0.3567	0.6433	0.1272	0.8405	0.0323
	50–64	75	1.0000	0.4535	0.5465	0.2335	0.7478	0.0187
		80	1.0000	0.5325	0.4675	0.3386	0.6176	0.0438
		85	1.0000	0.5877	0.4123	0.3995	0.5645	0.0360
	> 65	75	1.0000	0.8810	0.1190	0.7419	0.2173	0.0408
		80	1.0000	0.8634	0.1366	0.7388	0.1905	0.0707
		85	1.0000	0.8639	0.1361	0.7409	0.1857	0.0734
F	12–24	75	1.0000	0.2342	0.7658	0.2423	0.2189	0.5388
		80	1.0000	0.1618	0.8382	0.1992	0.1829	0.6179
		85	1.0000	0.1187	0.8813	0.1772	0.1489	0.6739
	25–34	75	1.0000	1.0000	0.0000	0.8657	0.1297	0.0047
		80	1.0000	0.9939	0.0061	0.8539	0.1302	0.0159
		85	1.0000	0.8965	0.1035	0.7439	0.2334	0.0228
	35–49	75	1.0000	1.0000	0.0000	0.9148	0.0632	0.0220
		80	1.0000	1.0000	0.0000	0.9058	0.0733	0.0209
		85	1.0000	1.0000	0.0000	0.8878	0.1037	0.0085
	50–64	75	1.0000	1.0000	0.0000	0.9613	0.0341	0.0047
		80	1.0000	1.0000	0.0000	0.9563	0.0271	0.0165
		85	1.0000	1.0000	0.0000	0.9116	0.0666	0.0218
	> 65	75	1.0000	1.0000	0.0000	0.9494	0.0303	0.0203
		80	1.0000	1.0000	0.0000	0.9681	0.0091	0.0228
		85	1.0000	1.0000	0.0000	0.9349	0.0273	0.0378

Table 2 (Continued).

Budget sizes	$k=1$	$k=1$	$k=2$	$k=1$	$k=2$	$k=3$
	1.0000	0.6228	0.3772	0.5181	0.3221	0.1599
Latent budgets	$k=1$	$k=1$	$k=2$	$k=1$	$k=2$	$k=3$
1. paidwork	0.0803	0.0351	0.1540	0.0000	0.2363	0.0265
2. dom.work	0.0779	0.1260	0.0000	0.1456	0.0000	0.0151
3. caring	0.0167	0.0238	0.0048	0.0244	0.0128	0.0000
4. shopping	0.0253	0.0331	0.0122	0.0358	0.0142	0.0138
5. per.need	0.0347	0.0360	0.0320	0.0370	0.0315	0.0332
6. eating	0.0623	0.0651	0.0568	0.0640	0.0687	0.0437
7. sleeping	0.3581	0.3554	0.3580	0.3637	0.3339	0.3891
8. educat.	0.0335	0.0000	0.0986	0.0000	0.0000	0.2094
9. particip	0.0146	0.0159	0.0122	0.0155	0.0168	0.0070
10. soc.cont	0.0656	0.0748	0.0497	0.0773	0.0568	0.0455
11. goingout	0.0297	0.0228	0.0406	0.0222	0.0327	0.0475
12. sports	0.0363	0.0388	0.0318	0.0426	0.0211	0.0467
13. gardening	0.0194	0.0216	0.0156	0.0198	0.0273	0.0021
14. outside	0.0061	0.0062	0.0058	0.0057	0.0083	0.0031
15. tv-radio	0.0779	0.0779	0.0768	0.0773	0.0830	0.0696
16. reading	0.0359	0.0413	0.0265	0.0417	0.0354	0.0181
17. relaxing	0.0070	0.0073	0.0063	0.0076	0.0059	0.0072
18. other	0.0188	0.0189	0.0184	0.0198	0.0154	0.0225

and the third latent component (characterized by education) is used predominantly by respondents younger than 25. Notice that the increase over time of the first latent component (domestic work) is again evident for males 25–64, but that this is at the expense of the second latent component (paid work) for 25–34 and 50–64, but not for 35–49. The principles of interpretation should be clear by now, and we leave the further interpretation of the solution to the reader.

6. Other algorithmic issues

Each step of the algorithm discussed in Sections 3 and 4 decreases the general least squares function f_x and therefore convergence to a minimum is ensured. We decide on convergence by comparing changes in subsequent changes in subsequent parameter estimates. When the difference between subsequent parameter estimates values is smaller than, e.g. 10^{-8}, we stop the algorithm.

The minimum found can be a local or a global minimum. Various sets of random starting values have to be chosen to be sure that a minimum found is indeed a global minimum. In order to investigate the seriousness of this local minimum problem, we have conducted studies on the data set in Table 1, and on some other data sets that have been analyzed previously with latent budget analysis. The data sets are the Manhattan Attitude data (first published by Srole et al. (1962) and analyzed with latent budget analysis by van der Heijden et al. (1989), the Amazon Indian data

(Gross et al. (1985); analyzed with latent budget analysis by de Leeuw et al. (1990), Aitchison's Data 10 (see Table 1), and the Dutch Shoplifting data (Israels, 1987) analyzed with latent budget analysis by van der Heijden et al. (1989). Since this local optimum problem is also known to exist for the maximum likelihood estimation procedure of the latent budget model, we have also performed simulations for this estimation technique. The simulation study was designed to evaluate the following properties:

- convergence properties, namely can we expect local optima and what stopping criterion is appropriate for the two algorithms;
- computation time of both algorithms;

In the simulation study analyzed the 5 data sets and we varied:

1. The starting values. 200 different sets of starting values were used for each data set;
2. The estimation procedure. In each analysis we obtained weighted least squares estimates and maximum likelihood estimates;
3. The number of latent budgets. The simulation study was carried out for the latent budget model with 2, 3 and 4 latent budgets, for each data set;
4. The convergence criterion. The simulation study was carried out with

- convergence on the parameter estimates, the procedure stops when the maximum difference between two parameter estimates in subsequent iterations is less than 10^{-5};
- convergence on the parameter estimates, the procedure stops when the maximum difference between two parameter estimates in subsequent iterations is less than 10^{-10};
- convergence on the function value, the procedure stops when the difference between the function values in two subsequent iterations is less than 10^{-14}.

So we conducted 5 (data sets) $\times$200 (starting values) $\times$2 (estimation procedures) $\times$3 (models) $\times$3 (convergence criteria)=18 000 analyses on a Pentium 200 mHz personal computer, characterizing 18 different analysis situations (cells) for each data set. For each cell we evaluated the mean computation time, and the number of local optima plus their deviance form the global optimum.

The results are displayed in Table 3(local optima) and Table 4(computation time). For $K = 2$, $K = 3$, and $K = 4$ we compare the results between maximum likelihood estimates and least squares estimates, with different convergence criteria. Convergence on the likelihood/loss function is included to compare the optima from the other analyses to a criterion that has fully converged.

In the top half of Table 3 the number on the left in each cell (indicated by #) denotes the number of optima obtained in the analysis, hence a 1 means that the algorithm converged to the same maximum every time; and the number on the right (indicated by "tot") denotes how many times the best function value, assumed to be the global optimum was obtained in 200 analyses. If only one optimum was obtained the number on the right automatically becomes 200. In Table 3 a function value is regarded as a separate optimum if the difference with other optima is greater than 10^{-8}. However, in the cases with local optima, not all differences between separate

Table 3
Results on convergence properties in the simulation study

		wls estimates						ml estimates					
		Parameters Func.						Parameters Func.					
		10^{-5}		10^{-10}		10^{-14}		10^{-5}		10^{-10}		10^{-14}	
Data set	K	#	tot	#	tot	#	tot	#	tot	#	tot	#	tot
Manh.	2	1	200	1	200	1	200	1	200	1	200	1	200
Attitude	3	1	200	1	200	1	200	1	200	1	200	1	200
(6 × 4)	4[a]												
Israel.	2	1	200	1	200	1	200	1	200	1	200	1	200
Worries	3	1	200	1	200	1	200	1	200	1	200	1	200
(5 × 9)	4	1	200	1	200	1	200	21	177	19	178	16	179
Aitchis.	2	1	200	1	200	1	200	1	200	1	200	1	200
Data 10	3	1	200	1	200	1	200	1	200	1	200	1	200
(20 × 6)	4	1	200	1	200	1	200	1	200	1	200	1	200
Amzon	2	1	200	1	200	1	200	1	200	1	200	1	200
Indians	3	1	200	1	200	1	200	10	190	5	192	6	191
(12 × 6)	4	2	121	2	121	2	121	139	1	10	43	69	26
Dutch	2	1	200	1	200	1	200	1	200	1	200	1	200
Shopl'g.	3	2	160	2	160	2	160	132	6	43	39	14	39
(18 × 13)	4	2	163	2	163	2	163	151	1	92	1	73	18
Worries	4	1	200	1	200	1	200	1	200	1	200	1	200
Amazon	3	1	200	1	200	1	200	6	190	4	192	5	191
Indians	4	2	121	2	121	2	121	8	146	8	148	9	149
Dutch	3	2	160	2	160	2	160	4	136	5	144	5	147
Shopl'g	4	2	163	2	163	2	163	5	151	5	151	5	151

[a]Not applicable, saturated model.

optima are significantly large. It was hoped for that in the cells where the convergence criterion$=10^{-14}$ the algorithm could truly converge, but the model did not converge fully for the maximum likelihood estimates of the Israelian Worries data ($K=4$), the Amazon Indian data ($K=3$, $K=4$) and the Shoplifting ($K=3$, $K=4$) data. Therefore the obtained optima need a closer look, to determine whether we have true local optima or just local optima due to a criterion that is too strict. If we regard a function value as a separate optimum if the difference with other function values is greater than 1, we find the results in the bottom half of Table 3.

For both definitions of a separate optimum (10^{-8} as well as 1) we see that the WLS algorithm (i.e. the active constraints method) is always superior to the EM algorithm. The WLS algorithm encounters less local optima, and the WLS algorithm is always fully converged at the convergence criterion 10^{-5}, while the EM-algorithm is only fully converged at the convergence criterion 10^{-5} for ($K=2$). For the larger data sets with $K>2$ convergence was sometimes not even obtained at the convergence criterion 10^{-14} although the differences were small between most of the values.

Table 4
Results on computation time in the simulation study

		wls estimates Parameters				ml estimates Parameters			
		10^{-5}		10^{-10}		10^{-5}		10^{-10}	
Data set	K	sec.	iter	sec.	iter	sec.	iter	sec.	iter
Manh.	2	0.05	8	0.06	9	0.16	184	0.18	246
Attitude	3	0.10	13	0.11	16	1.59	1856	2.03	2672
(6 × 4)	4[a]								
Israel.	2	0.10	15	0.10	18	0.18	159	0.20	204
Worries	3[b]	1.47	92	2.22	136	0.71	585	0.92	840
(5 × 9)	3[b]	0.15	12	0.16	15				
	4[c]	24.05	673	34.70	952	54.78	31 682	84.40	47 489
Aitchis.	2	0.19	17	0.20	21	1.76	615	1.63	851
Data 10	3	0.63	45	0.74	59	5.20	1867	6.52	2628
(20 × 6)	4	2.10	76	1.82	99	9.13	2595	11.88	3647
Amazon	2	0.31	32	0.37	40	0.35	218	0.53	286
Indian	3	3.04	141	4.40	209	1.71	940	2.20	1260
(12 × 6)	4	137.53	2594	179.20	3139	78.77	33 837	133.48	58 853
Dutch	2	1.18	45	1.46	57	0.45	109	0.64	140
Shopl'g	3	29.56	371	42.70	513	26.06	5503	40.13	9160
(19 × 9)	4	257.48	2261	403.17	3336	138.29	22 515	259.26	53 064

[a]Not applicable, saturated model.
[b]There are 3 outliers in the latent budget analysis of the Israelian Worries Data with $K=3$ and WLS parameter estimates. If these outliers are removed, the second line is applicable.
[c]The range of computation time for WLS is from 0.21 s to 13 min per analysis, for ML from 1.1 s to 1.5 min per analysis.

In Table 4 the results on computation time are presented. In each cell the mean number of seconds per analysis and the mean number of iterations per analysis are presented.

From Table 4 we can see that for the "Manhattan Attitude" data, the "Israelian Worry" data and "Aitchison's Data 10" the least squares algorithm is faster than the maximum likelihood algorithm. For the "Amazone Indian" data and the "Dutch Shoplifting" data the maximum likelihood algorithm is faster, if we compare between equal convergence criteria. However, from Table 3 we concluded that for weighted least squares estimates a convergence criterion 10^{-5} is appropriate, while for maximum likelihood we need at least 10^{-10}, and therefore we should compare the computation time of weighted least squares estimates with a convergence criterion 10^{-5} with maximum likelihood estimates with a convergence criterion 10^{-10}. In this case the weighted least squares algorithm and the maximum likelihood algorithm have approximately the same speed for the "Amazone Indian" data and the "Dutch Shoplifting" data.

7. Concluding remarks

In this paper we propose a (weighted) least squares algorithm for the estimation of a mixture model for compositional data. Previously, an algorithm inspired by least squares existed (see Renner, 1993a,b) where the observed matrix was approximated by another matrix, but not always in a least squares sense (see Section 1). A maximum likelihood algorithm exists that can be used when the data are realisations from a product-multinomial distribution. If this assumption holds, then this maximum likelihood algorithm is preferable to our (weighted) least squares algorithm because of the general properties of maximum likelihood estimation, such as efficiency of estimates. However, in many situations this assumption is not realistic, and in this situation the least squares algorithm can be applied.

A drawback of the least squares algorithm is that the asymptotic distribution of the function value under some null-hypothesis is unknown. We are currently studying whether the parametric bootstrap can be used to test for the number of latent compositions (compare Aitkin et al. (1981), who use this approach to test for the number of latent classes in latent class analysis). Secondly, we have not yet found a way to derive an asymptotic covariance matrix for the parameters, but we currently investigate whether it is possible to tackle this problem using either the parametric or the non-parametric bootstrap.

8. For further reading

Clogg, 1981, Gifi, 1990, Guttman, 1971, Heuer, 1979, Israëls, 1987, van der Ark and van der Heijden, 1998a,

Appendix A.

In this Appendix we give an algorithm for the non-negative least squares problem with equality restrictions. The solution of the $(n \times 1)$ column vector $\boldsymbol{x}$ from the least square problem $\mathrm{SSQ}(\boldsymbol{Qx} - \boldsymbol{r}) = \min$ with side condition $\boldsymbol{Cx} = \boldsymbol{d}$, will be written as $\boldsymbol{x}{:=}\mathrm{LSEQ}(\boldsymbol{Q}, \boldsymbol{r}, \boldsymbol{C}, \boldsymbol{d})$ and can be solved from the Eqs. (8), (9a) and (9b). $\boldsymbol{C}_1$ and $\boldsymbol{d}_1$ are defined in (4b). In the algorithm given below P and Z are sets consisting of sub-sets of the first n integers. $n(Z)$ is the number of elements of the set $Z, \boldsymbol{0}_{n(Z)}$ is a column vector with $n(Z)$ zero elements, and $\boldsymbol{E}_Z$ is a $(n(Z) \times n)$ matrix with rows equal to the row numbers given in Z of the identity matrix $\boldsymbol{I}_n$. In step 2 $\boldsymbol{y}$ is partitioned in $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$, where $\boldsymbol{y}_1$ corresponds to the equality restrictions $\boldsymbol{C}_1\boldsymbol{x} = \boldsymbol{d}_1$, and $\boldsymbol{y}_2$ corresponds to the equality restrictions $x_i = 0$. This algorithm is an adaptation of the NNLS algorithm of Lawson and Hanson (1995), p. 161.

Step 0: Set $\boldsymbol{C} = \boldsymbol{C}_1$ and $\boldsymbol{d} = \boldsymbol{d}_1$ and find $\boldsymbol{x}{:=}\mathrm{LSEQ}(\boldsymbol{Q}, \boldsymbol{r}, \boldsymbol{C}, \boldsymbol{d})$.

Step 1: Set $P = \{j : x_j > 0\}$, $Z = \{j : x_j \leq 0\}$; go to Step 12 if $Z = \{\varnothing\}$, otherwise go to Step 6.

Step 2: $\boldsymbol{y}{:=}(\boldsymbol{CC}')^{-1}\boldsymbol{CQ}'(\boldsymbol{r} - \boldsymbol{Qx}), \boldsymbol{w}(P) = \boldsymbol{0}_{n(P)}$ and $\boldsymbol{w}(Z){:=}\boldsymbol{y}_2$.

Step 3: If $Z = \{\emptyset\}$ or if $w_j \leq 0$ for all $j \in Z$, go to Step 12.
Step 4: Find an index $t \in Z$ such that $w_t = \max\{w_j : j \in Z\}$.
Step 5: Move the index t from set Z to set P.
Step 6: Define $\boldsymbol{C} = (\boldsymbol{C}_1/\boldsymbol{E}_Z)$ and $\boldsymbol{d} = (\boldsymbol{d}_1/\boldsymbol{0}_{n(Z)})$ and find $\boldsymbol{z}$:=LSEQ($\boldsymbol{Q}, \boldsymbol{r}, \boldsymbol{C}, \boldsymbol{d}$).
Step 7: if $z_j > 0$ for all $j \in P$, set $\boldsymbol{x}$:=$\boldsymbol{z}$ and go to Step 2.
Step 8: Find an index $q \in P$ such that $x_q/(x_q - z_q) = \min\{x_j/(x_j - z_j) : z_j \leq 0, j \in P\}$.
Step 9: Set $\alpha := x_q/(x_q - z_q)$.
Step 10: Set $\boldsymbol{x} := \boldsymbol{x} + \alpha(\boldsymbol{z} - \boldsymbol{x})$.
Step 11: Move from set P to set Z all indices $j \in P$ for which $x_j = 0$. Go to Step 6.
Step 12: *Comment*: The computation is completed.

Appendix B.

When we estimate the latent budget parameters by weighted least squares, then in Eq. (8) $\boldsymbol{r} \equiv \text{vec}(\boldsymbol{W}\boldsymbol{P}'\boldsymbol{V})$ and $\boldsymbol{Q} \equiv \boldsymbol{V}\boldsymbol{A} \otimes \boldsymbol{W}$, where $\boldsymbol{V}$ and $\boldsymbol{W}$ are diagonal matrices. Hence Eq. (8) becomes

$$\begin{aligned} \boldsymbol{x}^{\mathrm{O}} &= ((\boldsymbol{V}\boldsymbol{A} \otimes \boldsymbol{W})'(\boldsymbol{V}\boldsymbol{A} \otimes \boldsymbol{W}))^{-1}(\boldsymbol{V}\boldsymbol{A} \otimes \boldsymbol{W})'\text{vec}(\boldsymbol{W}\boldsymbol{P}'\boldsymbol{V}) \\ &= ((\boldsymbol{A}'\boldsymbol{V}^2\boldsymbol{A})^{-1} \otimes \boldsymbol{W}^{-2})\text{vec}(\boldsymbol{W}^2\boldsymbol{P}'\boldsymbol{V}^2\boldsymbol{A}) \\ &= \text{vec}(\boldsymbol{P}'\boldsymbol{V}^2\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{V}^2\boldsymbol{A})^{-1}). \end{aligned}$$

and the Kronecker-product has dissappeared from the equation, which reduces the computation time enormously.

In general, it holds that $(\boldsymbol{Q}'\boldsymbol{Q})^{-1} = (\boldsymbol{A}'\boldsymbol{V}^2\boldsymbol{A})^{-1} \otimes \boldsymbol{W}^{-2}$, which means that the inverse of $\boldsymbol{Q}'\boldsymbol{Q}$ which is needed in Eqs. (9a) and (9b) can be computed easily by taking the inverse of a $(K \times K)$ matrix instead of a $(JK \times JK)$ matrix.

References

Aitchison, J., 1986. The Statistical Analysis of Compositional Data. Chapman and Hall, London.

Aitkin, M., Anderson, D., Hinde, J., 1981. Statistical modelling of data on teaching styles. J. R. Statist. Soc. Ser. A 144, 419–461.

Björck, Å., 1996. Numerical Methods for Least Square Problems. Society for Industrial and Applied Mathematics, Philadelphia.

Clogg, C.C., 1981. Latent structure models of mobility. Amer. J. Sociol. 86, 836–868.

Cutler, A., Breiman, L., 1994. Archetypal analysis. Technometrics 36, 338–347.

Cutler, A., 1993. A branch and bound algorithm for convex least squares. Commun. Statist. Simulation Comput. 22, 305–321.

de Leeuw, J., van der Heijden, P.G.M., 1988. The analysis of time-budgets with a latent time-budget model. In: E. Diday et al. (Eds.), Data Analysis and Informatics, vol. 5. North-Holland, Amsterdam, pp. 159–166.

de Leeuw, J., van der Heijden, P.G.M., 1991. Reduced rank models for contingency tables. Biometrika 78, 232–239.

de Leeuw, J., van der Heijden, P.G.M., Verboon, P., 1990. A latent time-budget model. Statistica Neerlandica 44, 1–22.

Gifi, A., 1990. Nonlinear Multivariate Analysis. Wiley, New York.

Gill, P.E., Murray, W., 1974. Numerical Methods for Constrained Optimization. Academic Press, London.

Gross, D.J., Rubin, J., Flowers, N.M., 1985. All in a day's time: random spot check data can describe the texture of a day's activities. Paper presented at the Annual Meeting of the American Association for the Advancement of Science, Los Angeles, May 29, 1995.

Guttman, L., 1971. Measurement as structural theory. Psychometrika 36, 329–347.

Heuer, J., 1979. Selbstmord bei Kinder und Jugendlichen [Suicide of children and youth]. Ernst Klett Verlag, Stuttgard.

Israëls, A., 1987. Eigenvalue Techniques for Qualitative Data. DSWO Press, Leiden.

Knulst, W.P., van Beek, P., 1990. Tijd komt met de jaren: onderzoek naar tegenstellingen en veranderingen in dagelijkse bezigheden van Nederlanders op basis van tijdbudget onderzoek (Sociale en culturele studies 14, Sociaal en Cultureel Planbureau). VUGA, Den Haag.

Lawson, C.L., Hanson, R.J., 1995. Solving Least Squares Problems. Society for Industrial and Applied Mathematics, Philadelphia.

Renner, R.M., 1993a. The resolution of a compositional data set into mixtures of fixed source compositions. Appl. Statist. 42, 615–631.

Renner, R.M., 1993b. A constrained least squares subroutine for adjusting negative estimated element concentrations to zero. Comput. Geosci. 19, 1351–1360.

Siciliano, R., van der Heijden, P.G.M., 1994. Simultaneous latent budget analysis of a set of two way tables with constant row sum data. Metron 53, 155–179.

Srole, L., Langner, T.S., Michael, S.T., Opler, M.K., Rennie, T.A.C., 1962. Mental Health in the Metropolis: The Midtown Manhattan Study. McGraw-Hill, New York.

van der Ark, L.A., van der Heijden, P.G.M., 1998a. (in press). On the identifiability in the latent budget model. J. Classification.

van der Ark, L.A., van der Heijden, P.G.M., 1998. Graphical display of latent budget analysis and latent class analysis, with special reference to correspondence analysis. In: J. Blasius, M. Greenacre (Eds.), Visualisation of Categorical Data. Academic Press, New York, pp. 489–508.

van der Heijden, P.G.M., Mooijaart, A., de Leeuw, J., 1989. Latent budget analysis. In: A. Decarli, B.J. Francis, R. Gilchrist, G.U.H. Seeber (Eds.), Statistical Modelling, Proceedings, Trento, 1989. Springer, Berlin.

van der Heijden, P.G.M., Mooijaart, A., de Leeuw, J., 1992. Constrained latent budget analysis. In: C.C. Clogg (Ed.), Sociological Methodology 1992. vol. 22. Basil Blackwell, Cambridge. pp. 279–320.

van der Heijden, P.G.M., 1994. Endmember analysis and latent budget analysis (letter to the editor). Appl. Statist. 43, 527–528.

Weltje, G.J., 1997. Endmember modelling of compositional data: numerical–statistical algorithms for solving the explicit mixing problem. Mathematical Geology 29, 503–549.

Zangwill, W.I., 1969. Nonlinear Programming, a Unified Approach. Prentice-Hall, London.