Journal of Classification 16:117-137 (1999)

On the Identifiability in the Latent Budget Model

L. Andries van der Ark

Utrecht University

Peter G. M. van der Heijden

Utrecht University

Dirk Sikkel

Tilburg University

Van der Ark's research was supported by the Netherlands Research Council, Grant 510-78-406. The authors would like to thank the three anonymous reviewers for their helpful and constructive reviews and Phipps Arabie for his editorial comments.

Authors' adresses: L. Andries van der Ark, Department of Methodology and Statistics, Faculty of Social sciences, Utrecht University, PO box 80140, 3508 TC, Utrecht, the Netherlands, email: A.vanderArk@fss.uu.nl; Peter G.M. van der Heijden, Department of Methodology and Statistics, Faculty of Social sciences, Utrecht University, PO box 80140, 3508 TC, Utrecht, the Netherlands, email: P.vanderHeijden@fss.uu.nl; Dirk Sikkel, Center for Economic Research, Tilburg University, PO box 90153, 5000 LE, the Netherlands, e-mail: Dirk@pi.net.

Abstract: The latent budget model (LBM) is a reduced rank model for the analysis of compositional data. In the social sciences the LBM was originally proposed for the analysis of two-way contingency tables. The rows of the table are modelled as a mixture of typical or latent distributions (budgets). The mixing parameters can be used to classify the rows into typical groups. In geology the model is used for the analysis of physical mixing processes, and in this context it is known as the *end-member model*.

A major drawback of the latent budget model is that, in general, the model is not identifiable, which complicates the interpretation of the model considerably. This paper studies the geometry and identifiability of the latent budget model. Knowledge of the geometric structure of the model is used to specify an appropriate criterion to identify the model. The results are illustrated by an empirical data set.

Keywords: Latent budget analysis; Correspondence analysis; Identification; Mixture models; Compositional data; Chi-squared distance.

1. Introduction

Compositional data are data collected in an $I \times J$ two-way table such that the elements of each row are proportions, summing up to 1. Such a row vector is called a budget, and its elements are conditional proportions and called components. Compositional data play an important role in many disciplines (see Aitchison 1986).

The property that the elements of the budgets add up to 1 yields specific problems for data analysis, such as the absence of an interpretable covariance structure, and the difficulty of parametric modelling (Aitchison 1986). Aitchison proposed to base analysis on the covariances of the log-ratios of the first $J-1$ components and the J-th component. These log-ratios are assumed to be realizations from a $J-1$-variate normal distribution. In this way we can circumvent the dependence due to the sum constraint, but the parameters are relatively difficult to interpret in terms of the original components and the focus on differences among the observed budgets disappears, as a result of the transformation of the data.

An alternative approach to solve the specific compositional data problems that does not have these disadvantages, is the latent budget model (LBM), a mixture model for compositional data. The LBM approximates the I observed budgets, $\mathbf{p}_i$ $(i = 1, \ldots, I)$, by a mixture of K $(K \leq \min(I,J))$ typical or latent budgets, β_k $(k=1, \ldots, K)$, yielding expected budgets denoted by π_i. Let $p_{j|i}$ $(j = 1, \ldots, J)$ be the components of $\mathbf{p}_i$, let $\pi_{j|i}$ $(j = 1, \ldots, J)$ be the components of π_i, let $\alpha_{k|i}$ $(i = 1, \ldots, I, k = 1, \ldots, K)$ be the mixing parameters, and let $\beta_{j|k}$ $(j = 1, \ldots, J)$ be the components of β_k. The latent budget model with K latent budgets, denoted by LBM(K), is

$$\pi_{j|i} = \sum_{k=1}^{K} \alpha_{k|i}\, \beta_{j|k} \tag{1}$$

The parameters in (1) are conditional probabilities, indicated by the subscript, and are subject to equality constraints

$$\sum_{j=1}^{J} \pi_{j|i} = \sum_{k=1}^{K} \alpha_{k|i} = \sum_{j=1}^{J} \beta_{j|k} = 1 \tag{2}$$

and inequality constraints

$$0 \le \pi_{j|i} \le 1,\ 0 \le \alpha_{k|i} \le 1,\ 0 \le \beta_{j|k} \le 1\ . \tag{3}$$

LBM(1) is the independence model, with $\alpha_{1|i} = 1$ and $\beta_{j|1} = \pi_j$, where π_j is the marginal probability for column j. Let p_{ij} be the unconditional observed proportions and let π_{ij} be the expected unconditional probabilities. For the marginal proportions of model (1) the following properties hold: $\Sigma_{j=1}^{J} \pi_{ij} \equiv \pi_i = \Sigma_{j=1}^{J} p_{ij} \equiv p_i$ and $\Sigma_{i=1}^{I} \pi_{ij} \equiv \pi_j = \Sigma_{i=1}^{I} p_{ij} \equiv p_j$ hold, where π_{ij} are the expected unconditional probabilities and p_{ij} are unconditional observed proportions. The relative importance of a latent budget β_k is expressed by $\pi_k=\Sigma_i\, \pi_i\, \alpha_{k|i}$.

The model was suggested by Goodman (1974a), and worked out in detail by Clogg (1981). The model is used to classify I budgets, which may represent persons, groups or objects, into a small number of latent budgets, consisting of prototypical characteristics of the sample. LBM(2) and LBM(3) can be visualized so that the differences among the budgets can be expressed in distances (see van der Ark and van der Heijden 1997). For an elaborate interpretation of the LBM we refer to van der Heijden, Mooijaart and de Leeuw (1992), and for the estimation of the parameters we refer to de Leeuw, van der Heijden and Verboon (1990) and Mooijaart and van der Heijden (1995). For an empirical comparison between the results from LBM and the approach proposed by Aitchison see de Leeuw *et al.* (1990) and for an overview on mixture models, see Everitt and Hand (1981), Titterington, Smith and Makov (1985), and Lindsay (1995). In geology the LBM is known as the endmember model (see Renner 1993).

As an example to illustrate the results we analyze data from a budget survey of meat and meat products, conducted by Stichting Telepanel, and presented in Table 2. Over a period of a year, respondents were asked weekly about their expenses with respect to meat and meat products, "pork," "beef," "horsemeat, veal and mutton" (H.V.M.), "mince," "processed meat" (Proc.) and "snacks and assorted barbecue/grill and fondue meat" (Other). These budgets are broken down by the respondents' attitudes to meat related topics. The categories are classifications of factor scores: low (–), medium (=) and high (+). The factor scores were obtained by three factor

Table 1: *Factors obtained by three factor analyses on general attitudes towards food, attitudes towards meat, and attitudes towards meat products.*

factor	abbreviation	description
General attitudes towards food		
health	H	respondent values individual health and environmental durability
cooking	C	respondent likes cooking
ready made	R	respondent likes ready made and frozen food
quality	Q	respondent is prepared to pay extra for good quality.
Attitudes towards meat		
traditional meat	T	respondent is a traditional Dutch meat eater, wants daily meat with fat.
naturalness	N	respondent wants natural meat, and is opposed to bio-industry.
selectivity	Se	respondent does not buy much meat but buys high quality.
Attitudes towards meat products		
skepticism	Sk	respondent has a skeptical attitude towards the prepreparation of meat products.
luxury	L	respondent considers meat products a luxury.
versatility	V	respondent uses meat products for many purposes.

Table 2: *Red meat consumption in the Netherlands*

		red meat						
Topic /attitude		pork	beef	h,v,m	mince	proc.	other	total
health +	(H+)	.3054	.3197	.0287	.1819	.1161	.0481	1.000
health =	(H=)	.3117	.2672	.0275	.2102	.1393	.0441	1.000
health −	(H−)	.2632	.3451	.0216	.2164	.1258	.0280	1.000
cooking +	(C+)	.2868	.3026	.0172	.2084	.1438	.0411	1.000
cooking =	(C=)	.2985	.2859	.0243	.2115	.1369	.0429	1.000
cooking −	(C−)	.3196	.2980	.0403	.1914	.1100	.0409	1.000
ready-made −	(R−)	.2999	.3330	.0334	.1970	.0956	.0412	1.000
ready-made =	(R=)	.3102	.2934	.0230	.2096	.1315	.0322	1.000
ready-made +	(R+)	.2834	.2471	.0255	.2065	.1755	.0620	1.000
quality +	(Q+)	.2845	.3223	.0317	.2015	.1200	.0400	1.000
quality =	(Q=)	.2901	.2797	.0241	.2276	.1400	.0384	1.000
quality −	(Q−)	.3328	.2653	.0214	.1922	.1411	.0472	1.000
trad. meat +	(T+)	.2645	.2900	.0339	.2356	.1419	.0341	1.000
trad. meat =	(T=)	.3041	.2897	.0261	.2069	.1361	.0371	1.000
trad. meat −	(T−)	.3172	.3033	.0232	.1842	.1176	.0545	1.000
naturalness +	(N+)	.2891	.2520	.0313	.2242	.1521	.0514	1.000
naturalness =	(N=)	.3114	.2897	.0277	.2030	.1287	.0397	1.000
naturalness −	(N−)	.2997	.3320	.0217	.1923	.1180	.0364	1.000
selectivity +	(Se+)	.3126	.2718	.0285	.1951	.1352	.0568	1.000
selectivity =	(Se=)	.3032	.3124	.0261	.2042	.1210	.0331	1.000
selectivity −	(Se−)	.2869	.2906	.0255	.2162	.1418	.0389	1.000
scepticism +	(Sk+)	.3007	.2894	.0289	.1971	.1287	.0553	1.000
scepticism =	(Sk=)	.3039	.2800	.0275	.2080	.1384	.0423	1.000
scepticism −	(Sk−)	.2933	.3337	.0221	.2055	.1167	.0286	1.000
luxury +	(L+)	.3305	.3020	.0233	.1899	.1062	.0480	1.000
luxury =	(L=)	.2875	.3027	.0299	.2141	.1336	.0323	1.000
luxury −	(L−)	.2910	.2740	.0252	.2078	.1531	.0489	1.000
versatility +	(V+)	.3112	.3174	.0258	.2026	.1064	.0366	1.000
versatility =	(V=)	.3030	.2676	.0248	.2175	.1436	.0435	1.000
versatility −	(V−)	.2785	.3156	.0321	.1810	.1461	.0467	1.000

analyses on general attitudes towards food, attitudes towards meat, and attitudes towards meat products. The factors are given in Table 1.

Each subject is classified into one of the three categories of the above factors. Each individual budget is, therefore, used 10 times to calculate a group budget. The general idea behind this classification is to find latent budgets that not only show differences in purchase behavior, but also

Table 3: *Unidentified LBM(1), LBM(2) and LBM(3).*

		LBM(1)	LBM(2)		LBM(3)		
mixing parameters		$k=1$	$k=1$	$k=2$	$k=1$	$k=2$	$k=3$
health +	(H+)	1.000	.3418	.6582	.3979	.1472	.4549
health =	(H=)	1.000	.5059	.4941	.4152	.3526	.2321
health −	(H−)	1.000	.3100	.6900	.0960	.1383	.7657
cooking +	(C+)	1.000	.4305	.5695	.2588	.2737	.4675
cooking =	(C=)	1.000	.4609	.5391	.3346	.3042	.3612
cooking −	(C−)	1.000	.3846	.6154	.4687	.1939	.3374
ready-made −	(R−)	1.000	.2899	.7101	.3457	.0870	.5674
ready-made =	(R=)	1.000	.4312	.5688	.3673	.2632	.3695
ready-made +	(R+)	1.000	.6059	.3941	.3182	.4889	.1928
quality +	(Q+)	1.000	.3495	.6505	.2626	.1710	.5664
quality =	(Q=)	1.000	.4865	.5135	.2721	.3429	.3849
quality −	(Q−)	1.000	.5033	.4967	.5353	.3369	.1279
trad. meat +	(T+)	1.000	.4692	.5308	.1357	.3352	.5291
trad. meat =	(T=)	1.000	.4464	.5536	.3547	.2838	.3616
trad. meat −	(T−)	1.000	.3844	.6156	.4686	.1937	.3378
naturalness +	(N+)	1.000	.5687	.4313	.3166	.4421	.2413
naturalness =	(N=)	1.000	.4350	.5650	.4021	.2644	.3335
naturalness −	(N−)	1.000	.3182	.6818	.3219	.1252	.5529
selectivity +	(Se+)	1.000	.4866	.5134	.4596	.3236	.2169
selectivity =	(Se=)	1.000	.3716	.6284	.3360	.1912	.4729
selectivity −	(Se−)	1.000	.4583	.5417	.2635	.3082	.4283
scepticism +	(Sk+)	1.000	.4388	.5612	.3890	.2706	.3404
scepticism =	(Sk=)	1.000	.4739	.5261	.3693	.3169	.3138
scepticism −	(Sk−)	1.000	.3168	.6832	.2625	.1297	.6078
luxury +	(L+)	1.000	.3723	.6277	.5184	.1732	.3084
luxury =	(L=)	1.000	.4161	.5839	.2550	.2558	.4893
luxury −	(L−)	1.000	.5101	.4899	.3175	.3681	.3144
versatility +	(V+)	1.000	.3410	.6590	.3838	.1476	.4687
versatility =	(V=)	1.000	.5149	.4851	.3613	.3695	.2692
versatility −	(V−)	1.000	.3914	.6086	.2642	.2237	.5121
latent budgets							
Pork		.2999	.2979	.3014	.4103	.2288	.2498
Beef		.2960	.1204	.4260	.3056	.1481	.3841
Horsemeat, Veal, Mutton		.0268	.0258	.0275	.0282	.0251	.0266
Mince		.2042	.2473	.1723	.1336	.2835	.2143
Processed meat		.1307	.2356	.0530	.0609	.2594	.1077
Other		.0424	.0729	.0198	.0614	.0551	.0175

correspond to differences in attitudes that are relevant to marketing. We estimated the parameters of LBM(1), LBM(2) and LBM(3), by weighted least-squares (see Mooijaart and van der Heijden 1995). The results are presented in Table 3.

A major problem is that, in general, LBM(K) is not identifiable, which complicates the interpretation of the model considerably, and therefore also the classification of the budgets. LBM(2) and LBM(3) of Table 3 are not identifiable, and we cannot interpret them since parameter estimates with values completely different from those in Table 3 may yield exactly the same goodness of fit statistic. The unidentifiability can be demonstrated by writing the model in matrix notation. If we collect $\pi_{j|i}$ in an $I \times J$ matrix Π, $\alpha_{k|i}$ in an $I \times K$ matrix $\mathbf{A}$, and $\beta_{j|k}$ in a $J \times K$ matrix $\mathbf{B}$, then (1) can be written as $\Pi = \mathbf{A}\mathbf{B}^T$ from which it is clear that

$$\Pi = \mathbf{A}\mathbf{B}^T = \mathbf{A}\mathbf{T}^{-1}\mathbf{T}\mathbf{B}^T = \mathbf{A}^*\mathbf{B}^{*T} \qquad (4)$$

where $\mathbf{A}^* = \mathbf{AT}^{-1}$ and $\mathbf{B}^{*T} = \mathbf{TB}^T$ (see de Leeuw *et al.* 1990). $\mathbf{T}$ is a matrix of order $K \times K$ with elements τ_{cd}. To ensure that $\mathbf{A}^*$ and $\mathbf{B}^*$ satisfy the equality constraints in (2), de Leeuw *et al.* (1990) prove that $\mathbf{T}$ is subject to the constraint

$$\sum_{d}^{K} \tau_{cd} = 1 \quad (c = 1, \ldots, K). \tag{5}$$

Nevertheless, (5) still leaves an infinite number of possible values for τ_{cd}, with the restriction that (3) should hold.

In special cases, however, the model is identifiable. LBM(1) is identifiable since $\mathbf{T}$ is a 1×1 matrix that can only have the value 1 to meet the constraints in (2) and (3). Bandeen-Roche (1994) describes a model, equivalent to the LBM, on which she places distributional assumptions to ensure identifiability. Identifiability can also be achieved by constraining the parameters (see de Leeuw *et al.* 1990; van der Heijden *et al.* 1992; Goodman 1974b, for multiple latent class analysis). This is what is usually done. It is well known that for LBM(2) an identified solution is obtained by setting two parameters equal to zero (see Clogg 1981). Such fixed value constraints may decrease the model fit.

In this paper we propose a different method to identify the parameters of the model, emphasizing the explorative nature of the model. Instead of fixing parameters, we choose $\mathbf{T}$ such that the solution is optimal in a specific sense. This method is somewhat similar to the rotation problem in factor analysis, where unidentifiable solutions are usually varimax-rotated or oblimin-rotated to simplify the interpretation. We will call the common factor model identifiable because the varimax-rotated solution is always unique in practical situations. Likewise in this paper, unique solutions obtained by optimization of some criterion will also be called identified solutions. In Section 2, we discuss the geometry of the LBM and we make explicit how a feasible latent budget solution should be chosen. Within this framework we discuss several ways to identify the parameters in Section 3. De Leeuw *et al.* (1990) proposed identifiable solutions for LBM(2). We will discuss these solutions in our framework and extend their ideas to models of higher rank, for which there is so far no proposal for identifiability. In Section 4 the results are related to to correspondence analysis (CA).

2. Geometry in the Latent Budget Model

We will first discuss the geometry of LBM(2), then of LBM(3), and finally we will extend the discussion to LBM(K).

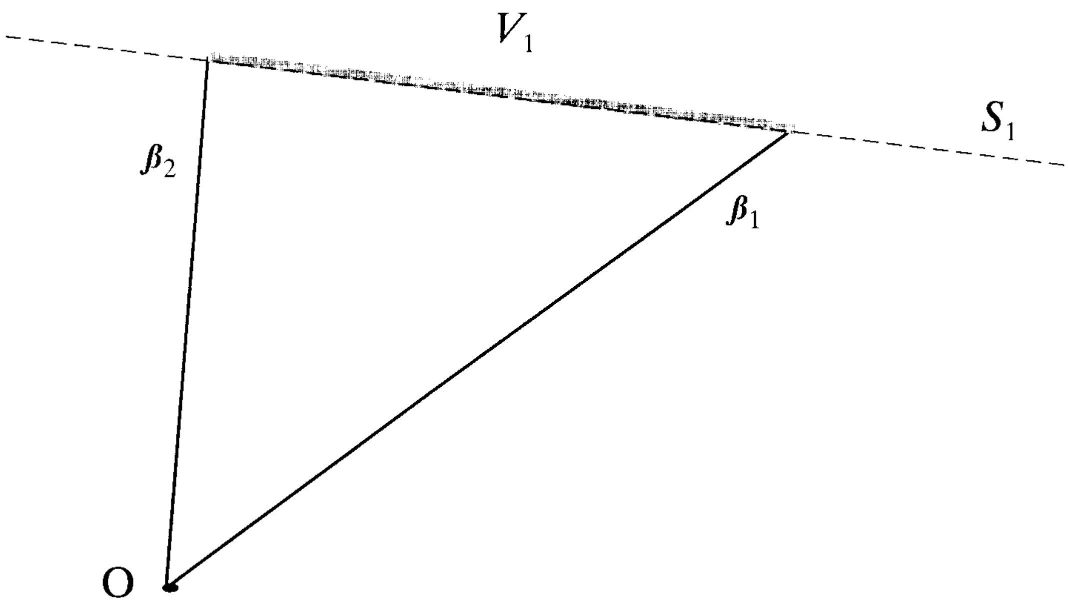

Figure 1a. The space spanned by two latent budgets and the space spanned by three latent budgets.

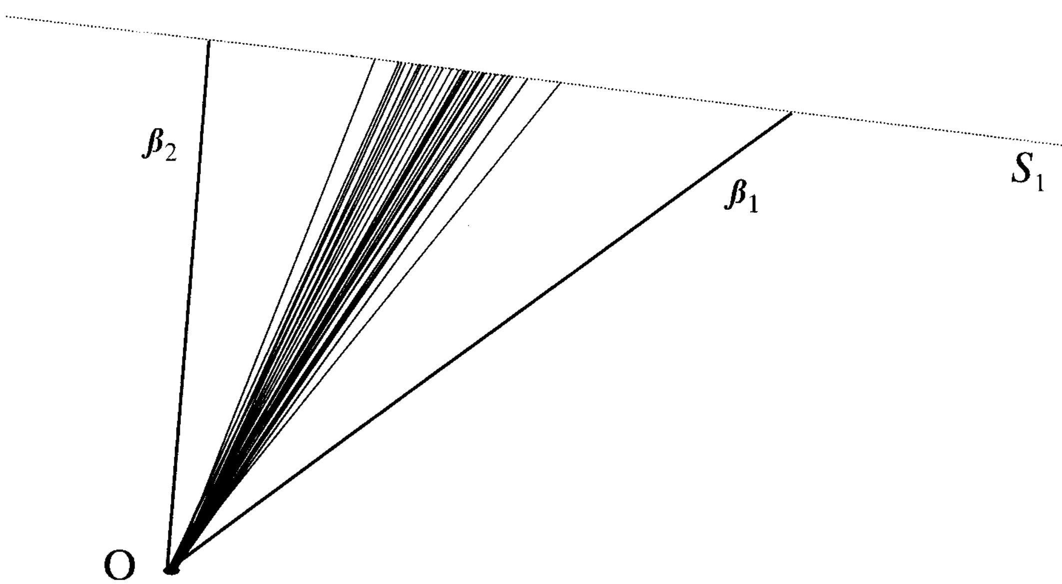

Figure 1b. The expected budgets as a combination of the latent budgets.

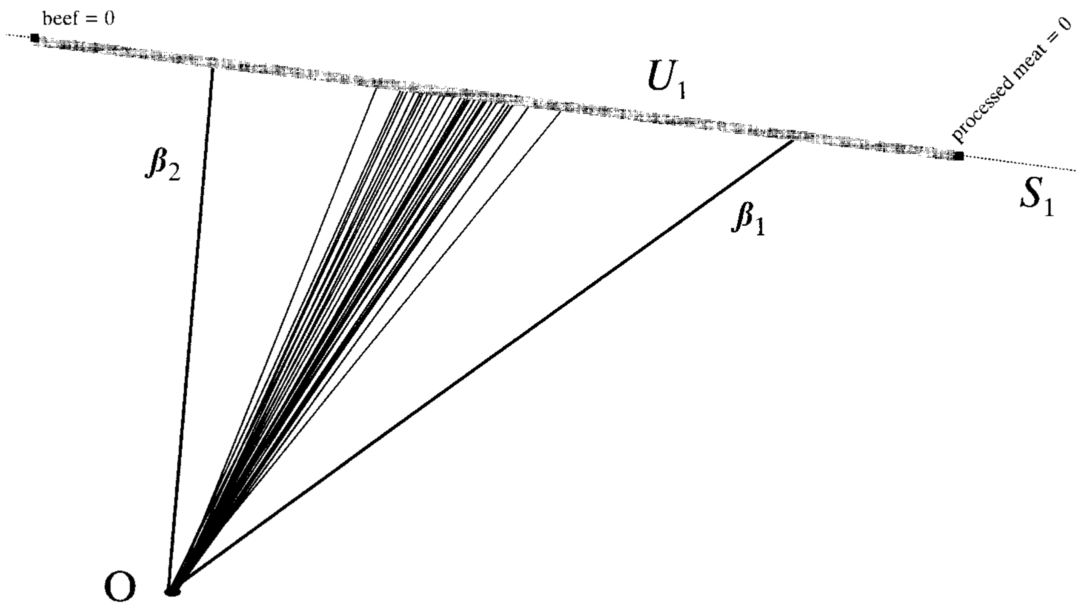

Figure 1c. U_1, the subspace of bugets.

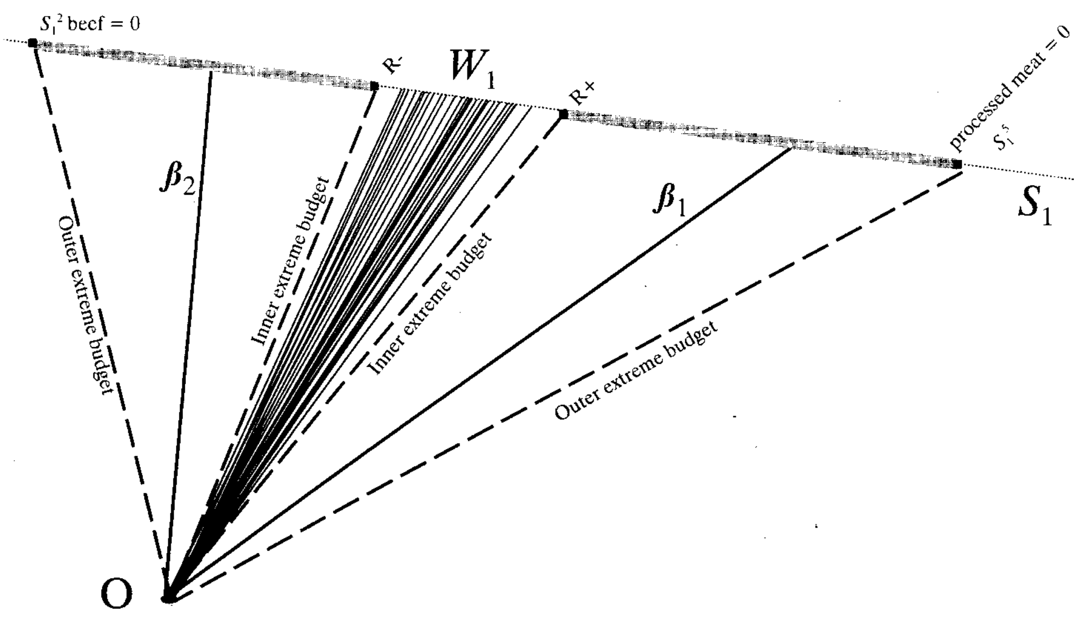

Figure 1d. Graphic display of the outer extreme budgets and the inner extreme budgets of LBM(2).

In Figure 1 graphic displays of LBM(2) are given. In LBM(2), the unidentified latent budgets β_1 and β_2 can be viewed as two vectors in J-dimensional space. The heads of any two vectors can be connected by a line segment, denoted by V_1, which is a subset of the line S_1 (see Figure 1a). The subscript '1' denotes the dimensionality of the space.

The expected budgets $\pi_1, \ldots, \pi_I$ are J-dimensional vectors and convex combinations of β_1 and β_2. Therefore the heads of $\pi_1, \ldots, \pi_I$ lie on V_1, and the relative distance from $\pi_1, \ldots, \pi_I$ to β_1 and β_2 is expressed by the mixing parameters. A graphic representation of the parameter estimates of LBM(2) in Table 3 is presented in Figure 1b.

The unidentified latent budgets β_1 and β_2, collected in $\mathbf{B}$, can be transformed into $\mathbf{B}^*$ such that the identified latent budgets β_1^* and β_2^* lie on S_1, by $\mathbf{B}^{*T} = \mathbf{T}\mathbf{B}^T$ (see 4), with τ_{cd} subject to the constraints in (5). Let $\mathbf{b}$ be a vector such that $\mathbf{b}^T = \tau_c\mathbf{B}^T$, where τ_c is the c-th row of $\mathbf{T}$. Hence $\mathbf{b} \in S_1$. If $\mathbf{b}$ has no negative elements then it is a budget.

If we choose τ_c such that one element of $\mathbf{b}$, say b_j, equals zero, then this point on S_1 is denoted by S_1^j. On one side of S_1^j there are vectors with $b_j > 0$, and on the other side vectors with $b_j < 0$. We can mark J such points on S_1, one for each element. In this way we can determine the region of S_1 which has vectors consisting solely of nonnegative elements, and thus are budgets. The region of budgets is denoted by U_1. A graphic representation is given in Figure 1c. The vectors that bound U_1 are called *outer extreme budgets*, and have one component equal to zero. LBM(2) always has two outer extreme budgets.

Not every $\mathbf{b} \in U_1$ is a feasible latent budget. A latent budget cannot lie between two expected budgets, because this would result in negative mixing parameters. Hence a latent budget cannot lie within the space spanned by the expected budgets that take the most extreme position on S_1. This space is denoted as W_1, and the most extreme expected budgets are called *inner extreme budgets*. For Table 3 the inner extreme budgets are the expected budgets of R+ and R−. LBM(2) always has two inner extreme budgets. A graphic representation of the inner extreme budgets, the outer extreme budgets, and the space where possible latent budgets may be positioned, $U_1 \setminus W_1$ (shaded area) is given in Figure 1d.

We turn now to the geometry in LBM(3). In LBM(3) the heads of the three unidentified latent budgets β_1, β_2 and β_3 span a triangular two dimensional plane, denoted by V_2, $V_2 \subset S_2$. A graphical representation is given in Figure 2.

The expected budgets are convex combinations of the latent budgets and, therefore, they lie in the triangle V_2. We represent S_2 in Figure 3 in two dimensions. In Figure 3 the upper vertex represents β_1, the right hand vertex represents β_2, and the left hand vertex represents β_3.

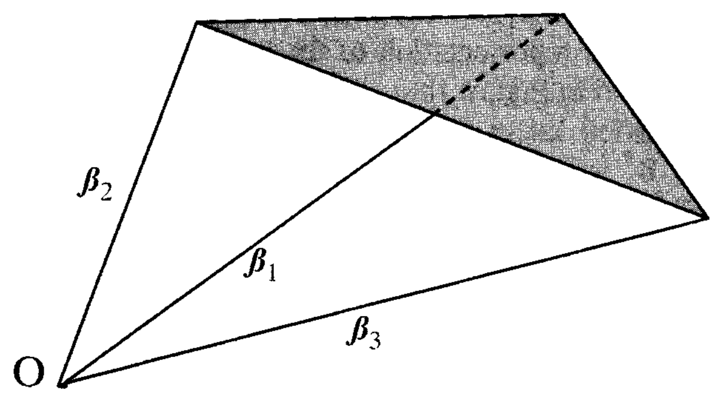

Figure 2. V_2. the space spanned by three unidentified latent budgets.

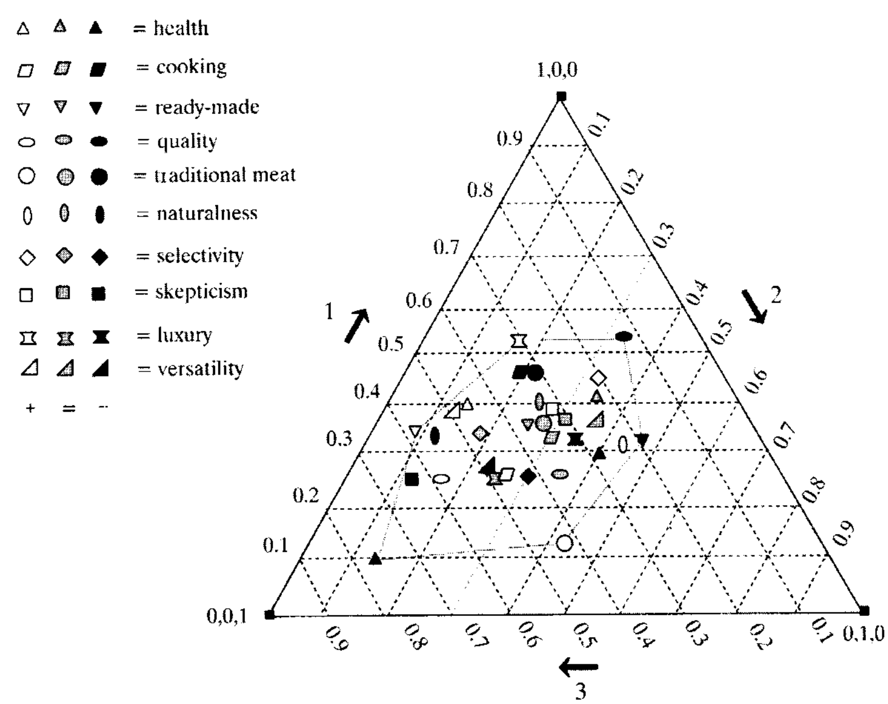

Figure 3. The expected budgets as a convex combination of β_1, β_2 and β_3. W_2, the space spanned by the inner extreme budgets.

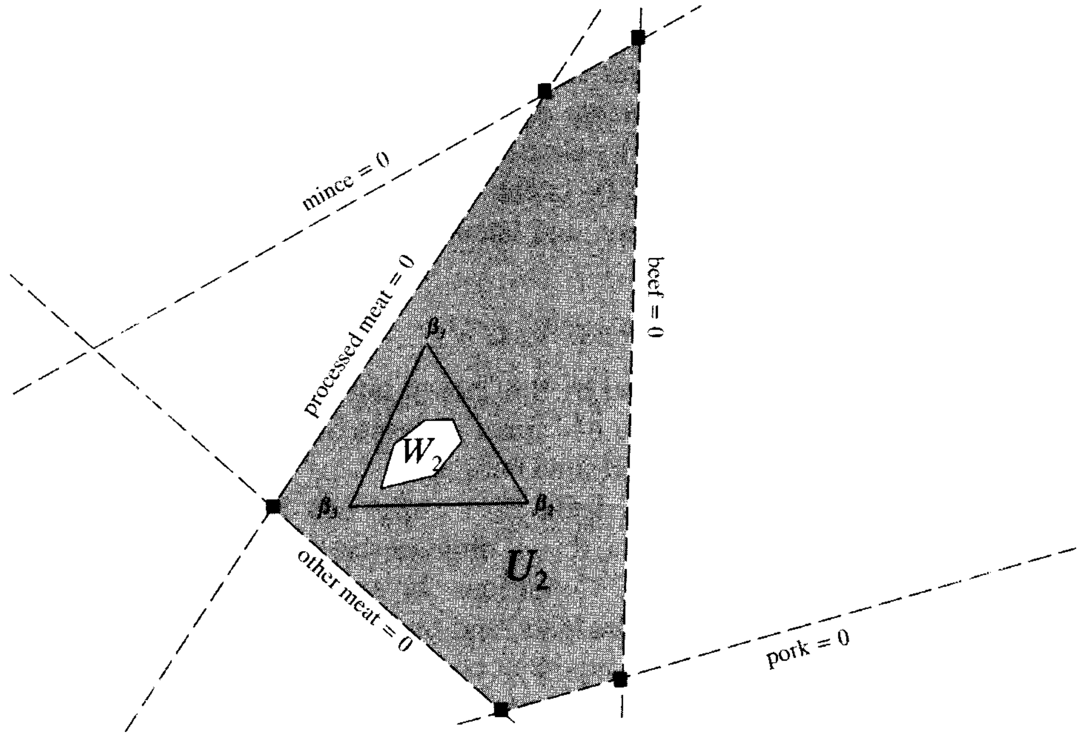

Figure 4. U_2, the space spanned by the outer extreme budgets, and the regions where **b** has one element equal to zero.

The positions of the expected budgets depicted in Figure 3 are determined by their mixing parameters (see Table 3 right panel), so the position of π_i is $\alpha_{1|i}$ times the distance from the bottom side to the upper vertex, $\alpha_{2|i}$ times the distance from the left hand side to the right hand vertex, and $\alpha_{3|i}$ times the distance from the right hand side to the left hand vertex. The expected budgets that have the most extreme position on S_2 are those of H–, R+, L+, Q–, R– and N+. These are the inner extreme budgets. LBM(3) may have more inner extreme budgets than latent budgets. The space spanned by the heads of the inner extreme budgets will be denoted as W_2.

Analoguous to LBM(2) there exist vectors $\mathbf{b} \in S_2$, where $\mathbf{b} = \tau_c \mathbf{B}^T$, that have one element equal to zero. In LBM(3) these vectors can be represented as lines on S_2, see Figure 4. The solid triangle represents V_2, and a dashed line represents the collection of vectors **b** with one element, say b_j, equal to zero. For b_j this line in S_2 is denoted by S_2^j and represented in Figure 4 by the label of the category that equals zero. S_2^j ($j = 1, \dots, J$) have vectors with $b_j > 0$ on one side and $b_j < 0$ on the other side. In this way we can determine the region of budgets, denoted by U_2. In Figure 4 S_2^1 (Pork = 0), S_2^2 (Beef = 0), S_2^4 (Mince = 0) and S_2^5 (Processed meat = 0) and S_2^6 (Other =

0) define the convex region U_2. The intersection of S_2^j and $S_2^{j'}$ represents a vector with both the j-th and the j'-th elements zero. The corner points of U_2 are the outer extreme budgets for LBM(3). Hence in the LBM(3) the outer extreme budgets have two elements equal to zero. It is relatively simple to determine the region of budgets by elementary row operations on **B**.

Analoguous of the LBM(2), latent budgets should lie in $U_2 \setminus W_2$ to avoid negative mixing parameters, and latent budgets should be chosen such that

$$W_2 \subseteq V_2^* \subseteq U_2 \tag{6}$$

where V_2^* is the subspace spanned by the chosen latent budgets. LBM(3) has a minimum of 3 and a maximum of I inner extreme budgets, and a minimum of 3 and a maximum of J outer extreme budgets.

In LBM(K) the latent budgets $\beta_1, \ldots, \beta_K$ can be viewed as K J-dimensional vectors that span a $K-1$ dimensional subspace V_{K-1} in the $(K-1)$-dimensional space S_{K-1}. Conditions for a sound latent budget solution require that both the latent budgets, and the mixing parameters meet the constraints in (2) and (3). The equality constraints in (2) are satisfied by imposing equality constraints on **T**, see (5) and in addition, the inequality constraints (3) are satisfied by extending (6) to the LBM(K),

$$W_{K-1} \subseteq V_{K-1}^* \subseteq U_{K-1} \tag{7}$$

where U_{K-1} is the space spanned by all budgets **b** that have $K-1$ elements equal to zero, and all other elements nonnegative; W_{K-1} is the space spanned by the inner extreme budgets; and V_{K-1}^* is a proposed identified latent budget solution.

3. Identifiability of the LBM by Optimization of a Criterion

In this section we will first discuss existing proposals for an identified LBM(2) and then we will use the geometric results of the previous section to work out ideas for an identified LBM of higher rank. For LBM(2), four typical identified solutions have already been indicated by de Leeuw *et al.* (1990). We label these

- the *outer extreme solution*, such that β_1^* and β_2^* are the two outer extreme budgets;
- the *inner extreme solution*, such that β_1^* and β_2^* are the two inner extreme budgets;
- the *emphasized first budget solution*, such that β_1^* is an inner extreme budget, and β_2^* is an outer extreme budget. Thus β_1^* is emphasized in the sense that for $k = 1$, π_k is as large as possible;

Table 4: *Four identified LBM(2) solutions*

		outer extreme solution		inner extreme solution		emphasized β_1 solution		emphasized β_2 solution	
mixing parameters		$k=1$	$k=2$	$k=1$	$k=2$	$k=1$	$k=2$	$k=1$	$k=2$
health +	(H+)	.3751	.6249	.1645	.8355	.7052	.2948	.0471	.9529
health =	(H=)	.4726	.5274	.6837	.3163	.8884	.1116	.1957	.8043
health −	(H−)	.3562	.6438	.0636	.9364	.6696	.3304	.0182	.9818
cooking +	(C+)	.4278	.5722	.4451	.5549	.8042	.1958	.1274	.8726
cooking =	(C=)	.4459	.5541	.5413	.4587	.8382	.1618	.1549	.8451
cooking −	(C−)	.4005	.5995	.2999	.7001	.7530	.2470	.0858	.9142
ready-made −	(R−)	.3443	.6557	.0000	1.000	.6472	.3528	.0000	1.000
ready-made =	(R=)	.4282	.5718	.4472	.5528	.8050	.1950	.1280	.8720
ready-made +	(R+)	.5319	.4681	1.000	.0000	1.000	.0000	.2862	.7138
quality +	(Q+)	.3797	.6203	.1888	.8112	.7138	.2862	.0540	.9460
quality =	(Q=)	.4610	.5390	.6222	.3778	.8667	.1333	.1781	.8219
quality −	(Q−)	.4710	.5290	.6755	.3245	.8855	.1145	.1933	.8067
trad. meat +	(T+)	.4507	.5493	.5673	.4327	.8473	.1527	.1624	.8376
trad. meat =	(T=)	.4372	.5628	.4954	.5046	.8220	.1780	.1418	.8582
trad. meat −	(T−)	.4004	.5996	.2993	.7007	.7528	.2472	.0857	.9143
naturalness +	(N+)	.5098	.4902	.8823	.1177	.9585	.0415	.2525	.7475
naturalness =	(N=)	.4304	.5696	.4592	.5408	.8092	.1908	.1314	.8686
naturalness −	(N−)	.3611	.6389	.0896	.9104	.6788	.3212	.0256	.9744
selectivity +	(Se+)	.4611	.5389	.6225	.3775	.8668	.1332	.1782	.8218
selectivity =	(Se=)	.3928	.6072	.2585	.7415	.7384	.2616	.0740	.9260
selectivity −	(Se−)	.4443	.5557	.5329	.4671	.8352	.1648	.1525	.8475
scepticism +	(Sk+)	.4327	.5673	.4713	.5287	.8135	.1865	.1349	.8651
scepticism =	(Sk=)	.4535	.5465	.5823	.4177	.8526	.1474	.1667	.8333
scepticism −	(Sk−)	.3603	.6397	.0854	.9146	.6773	.3227	.0244	.9756
luxury +	(L+)	.3932	.6068	.2609	.7391	.7392	.2608	.0747	.9253
luxury =	(L=)	.4192	.5808	.3993	.6007	.7881	.2119	.1143	.8857
luxury −	(L−)	.4751	.5249	.6970	.3030	.8931	.1069	.1995	.8005
versatility +	(V+)	.3746	.6254	.1618	.8382	.7042	.2958	.0463	.9537
versatility =	(V=)	.4779	.5221	.7120	.2880	.8984	.1016	.2038	.7962
versatility −	(V−)	.4046	.5954	.3213	.6787	.7605	.2395	.0920	.9080
latent budgets									
Pork		.2965	.3024	.2993	.3004	.2993	.3024	.2965	.3004
Beef		.0000	.5146	.2409	.3374	.2409	.5146	.0000	.3374
Horsemeat, Veal, Mutton		.0252	.0280	.0265	.0270	.0265	.0280	.0252	.0270
Mince		.2769	.1506	.2178	.1941	.2178	.1506	.2769	.1941
Processed meat		.3076	.0000	.1636	.1059	.1636	.0000	.3076	.1059
Other		.0938	.0044	.0520	.0352	.0520	.0044	.0938	.0352
budget proportions (π_k)		.4247	.5753	.4287	.5713	.7984	.2016	.1227	.8773

- the *emphasized second budget solution*, which is the opposite of the emphasized first budget solution. Now for $k = 2$, π_k is as large as possible.

The identified parameters of LBM(2) of the data in Table 2 are presented in Table 4. A graphic representation can be derived from Figure 1d in the following way. For the outer extreme solution we take the two outer extreme budgets as latent budgets; for the inner extreme solution we take the two inner extreme budgets as latent budgets; for the emphasized β_1 solution we take the outer extreme budget on the left hand side and the inner extreme budget on the right hand side as latent budgets; for the emphasized β_2 we take the opposites.

For the purpose of interpretation, the following arguments apply when choosing one of these options. In the outer extreme solution the latent budgets are as different as possible, and this will in most cases simplify their interpretation. Specifically, in each latent budget there is one component equal to zero, which will in many cases largely determine the interpretation. On the other hand, since the latent budgets are so extreme, the mixture coefficients of row i will be as close as possible to the mixture coefficients of row i', and this will make the interpretation of differences between rows more difficult.

The appropriate way to interpret the outer extreme solution in Table 4 is, therefore, as follows: $\hat{\beta}_1^*$ emphasizes the consumption of mince, processed meat and the category "Other," compared to the marginal budget, while $\hat{\beta}_2^*$ emphasizes the consumption of beef. There is no distinction between the budgets in the consumption of pork and horsemeat, veal and mutton. $\hat{\beta}_1^*$ can be described as a budget on low fat, but more expensive and elaborate meat, while $\hat{\beta}_2^*$ can be described as a budget on cheaper, faster, high fat meats.

In the inner extreme solution, the latent budgets are as similar as possible. At the same time, the mixture coefficients will be as different as possible. In particular, expected budgets of the rows that are the identified latent budgets have one mixture coefficient equal to zero and the other equal to 1. $\hat{\beta}_1^*$ of the inner extreme solution in Table 4 is interpreted in terms of the mixing parameters and corresponds to a greater emphasis on "ready-made" and "naturalness," and a not so much on "quality" and "skepticism." Since the mixing parameters that correspond to $\hat{\beta}_2^*$ are one minus the mixing parameters that correspond to $\hat{\beta}_1^*$, the interpretation of $\hat{\beta}_2^*$ is the opposite of the interpretation of $\hat{\beta}_1^*$. The attitude towards "ready-made" food is the most discriminating aspect of meat consumption.

In choosing either the emphasized first or second latent budget, the actual values of π_k, $(k = 1,2)$ could play a role. If one wanted to emphasize one of the latent budgets at the expense of the other, then latent budget k should be chosen for which π_k is largest. Emphasizing the first latent budget implies taking the first latent budget of the inner extreme solution and the second latent budget from the outer extreme solution.

In LBM(K) ($K > 2$), the identifiability problem is more complex than in LBM(2), due to the fact that there may be more than K inner extreme budgets and more than K outer extreme budgets (see Section 2). Therefore, we cannot extend the proposals for identifiability of LBM(2), to models of higher rank. For the identifiability of the LBM(3) certain proposals were introduced. De Leeuw *et al.* (1990) proposed to make as many parameters as possible equal to zero, a procedure which leaves open many alternatives. Ripley (1990) proposed to fit the model by placing two expected budgets on each boundary of V_2. Although he noticed that this does not always yield a feasible solution, he did not pursue the problem any further. Renner (1995) proposed choosing the latent budgets as close to the data as possible, but he did not use any explicit criterion. No proposal for identifiability for $K > 3$ so far exists.

We adopt the idea of inner extreme solutions and outer extreme solutions for LBM's of higher rank to facilitate model interpretation and do not consider emphasized latent budget solutions since they would become too numerous. In the LBM(2) the expressions "as similar as possible" and "as different as possible" are concepts that leave no room for interpretation, since all the budgets lie on a single line. For LBM's of higher rank there are a number of possibilities for defining these concepts explicitly.

1. *sums of distances among the identified latent budgets.* There are $\binom{K}{2}$ distances among the K latent budgets, and one option is to maximize the sum of these distances to obtain an outer extreme solution, and to minimize this sum to obtain an inner extreme solution. This still leaves open the choice of a specific distance measure. We propose using the chi-squared distance, δ_{χ^2} (see, for example, Greenacre 1984), a concept well-known in CA, rather than the Euclidean distances. In the chi-squared distance, distances between latent components are weighted by the average component size $1/\pi_j$, thus correcting for the fact that differences between $\beta_{j|k}$ and $\beta_{j|k'}$ are likely to be smaller for categories with a low marginal proportion π_j. The chi-squared distance between β_k and $\beta_{k'}$ is defined as
$$\delta_{\chi^2} = \left[\sum_{j=1}^{J} \frac{(\beta_{j|k} - \beta_{j|k'})^2}{\pi_j} \right]^{1/2}$$
2. *optimizing the volume of* V^*_{K-1}. In order to be able to calculate the hypervolume, distances between the identified latent budgets have to be known, and for these we could also use the chi-squared distances.
3. *measuring the dependence in* $\mathbf{B}^*$. For measuring the dependence in $\mathbf{B}^*$ it seems most elegant to weight for the number of observations falling in each of the latent budgets, that is, to work out the elements $\pi^*_{jk} = \pi^*_k \pi^*_{j|k}$. The dependence can then be measured using a criterion like the Pearson chi-square, the likelihood ratio chi-square, or some information criterion, which can be optimized to obtain an outer extreme and an inner extreme latent budget solution.

For mainly practical reasons we choose to identify the LBM by optimizing the sum of chi-squared distances. Optimizing the volume of V^*_{K-1} turned out to be too complicated to expand to higher rank models, and too time-consuming to use in an iterative procedure. The last criterion is easily expanded to LBM(K), but it is still not easy to optimize.

Hence, for an identified inner extreme solution and an identified outer extreme solution, we propose the criteria

$$\min\left[\sum_{q=1}^{Q}\delta_{\chi^2 q}\right] \quad \text{and} \quad \min\left[\frac{1}{\sum_{q=1}^{Q}\delta_{\chi^2 q}}\right]. \tag{8}$$

respectively, where $Q = \binom{K}{2}$.

We use a Monte Carlo procedure to compute solutions proposed in (8). We modified the well-known Metropolis Algorithm (Metropolis, Rosenbluth, Rosenbluth, Teller and Teller 1953; see also Press, Flannery, Teukolsky and Vetterling 1989, for computational issues). The central idea is as follows: we start with the unidentified latent budget solution, hence $\mathbf{T}_0 = \mathbf{I}_K$. Then a cycle of subprocesses starts, indexed by m $(m = 1, \ldots, M)$. Each subprocess consists of a number of iterations, indexed by n $(n = 1, \ldots, N)$. In each iteration a constant a_m is added to one of the elements of $\mathbf{T}$, and a_m is subtracted from another element chosen randomly from the same row, and the function in (8) is evaluated. a_m is fixed throughout the m-th subprocess. The probability of preferring the altered matrix at iteration n, $\mathbf{T}_{m,n}$ over $\mathbf{T}_{m,n-1}$ at iteration $n-1$ is

$$P\,(\text{accept}\ \mathbf{T}_{m,n}) = e^{\frac{f_{m,n}-f_{m,n-1}}{CV_n}}$$

in which $f_{m,n}$ is the function value (the sum of chi-squared distances at subprocess m and iteration n), and CV_n is a control variable that starts off at a high value in the first iteration and decreases after each iteration. Note that if $f_{m,n} > f_{m,n-1}$, then the probability is greater than 1 and the altered matrix $\mathbf{T}$ at iteration n is always accepted. The m-th subprocess stops when CV_n is sufficiently small and there is no more improvement in further minimizing (8). After the m-th subprocess, the constant a_m is diminished by a factor C, such that $a_{m+1} = a_m C$, and the $m+1$-th subprocess begins. The cycle of subprocesses stops when a_m is sufficiently small. The Metropolis algorithm is known for its ability to walk out of local optima, although a global optimum is not guaranteed and different altering schemes and different sets of unidentified parameters should be used as starting values to increase the possibility of finding a global optimum.

The outer extreme solution and the inner extreme solution of LBM(3) for the data in Table 2 are presented in Table 5. Graphic representations are given in Figure 5. The identified latent budgets in the outer extreme solution are three of the five outer extreme budgets. However, we stress that the identified latent budgets of the outer extreme solution need not be outer extreme budgets if $(K > 2)$.

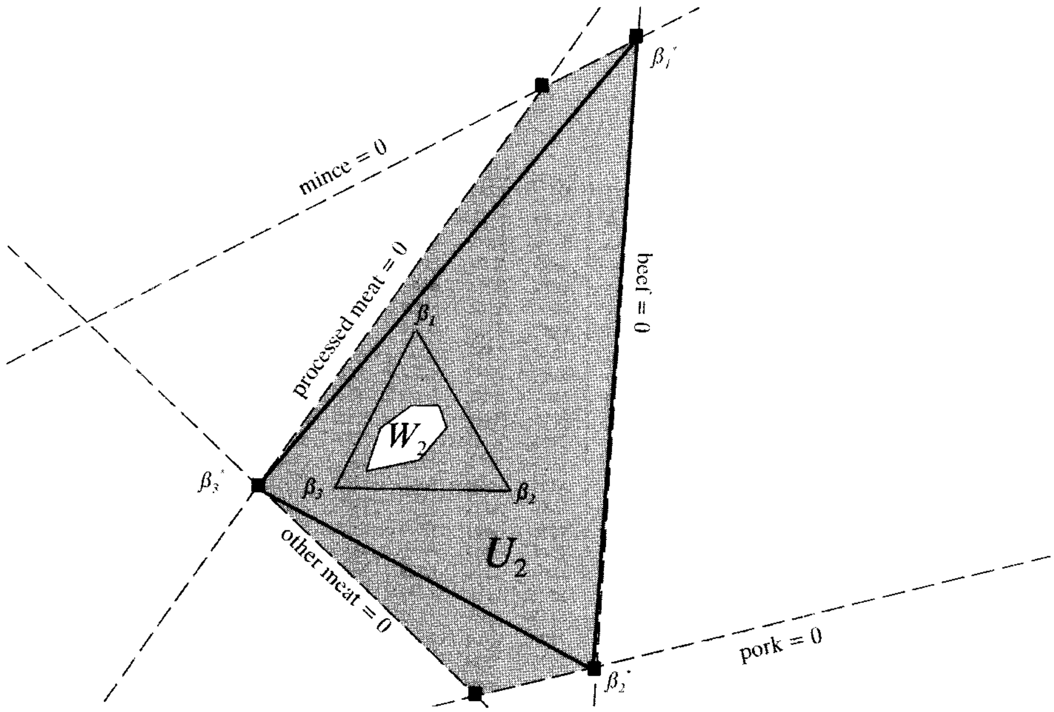

Figure 5a. Outer extreme solution in LBM(3).

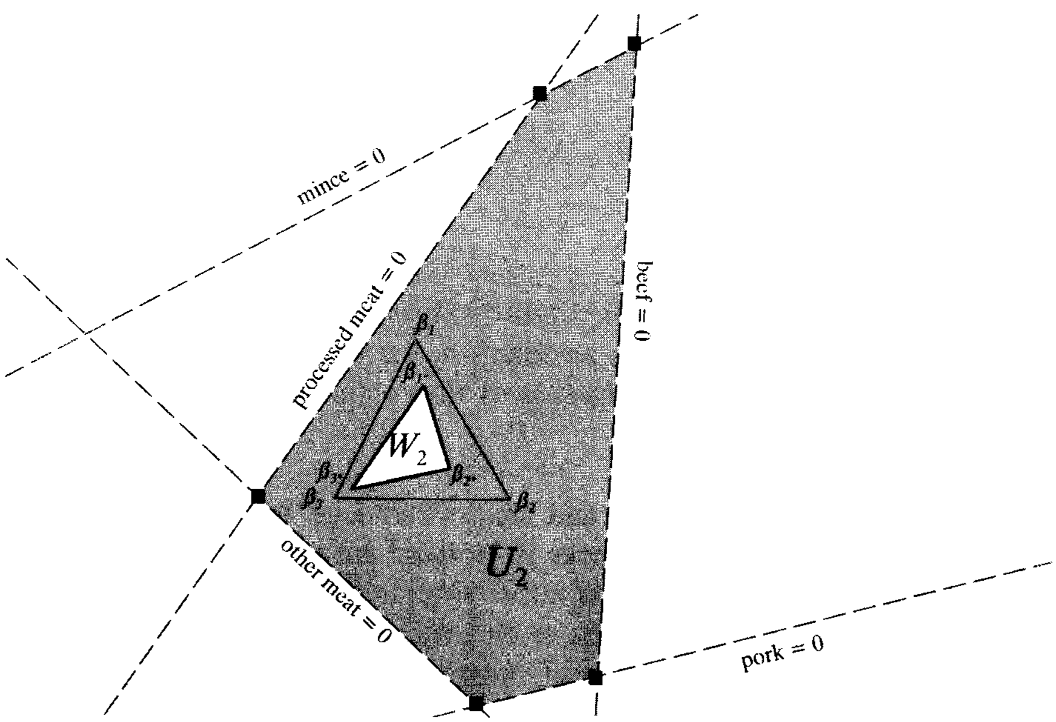

Figure 5b. Inner extreme solution in LBM(3).

Table 5: *Parameter estimates of the identified inner extreme solution and the identified outer extreme solution for LBM(3)*

		outer extreme solution			inner extreme solution		
mixing parameters		$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$
health +	(H+)	.1818	.2141	.6041	.4907	.0862	.4231
health =	(H=)	.2167	.2738	.5095	.4322	.4666	.1011
health −	(H−)	.0978	.2491	.6531	.0197	.2310	.7493
cooking +	(C+)	.1623	.2696	.5682	.2193	.4006	.3801
cooking =	(C=)	.1875	.2693	.5432	.3256	.4179	.2564
cooking −	(C−)	.2081	.2193	.5726	.5825	.1369	.2805
ready-made −	(R−)	.1587	.2024	.6389	.4336	.0000	.5664
ready-made =	(R=)	.1904	.2529	.5567	.3941	.3228	.2831
ready-made +	(R+)	.2101	.3270	.4629	.2228	.7772	.0000
quality +	(Q+)	.1482	.2381	.6136	.2681	.2037	.5282
quality =	(Q=)	.1761	.2888	.5352	.2113	.5248	.2638
quality −	(Q−)	.2473	.2541	.4987	.6275	.3725	.0000
trad. meat +	(T+)	.1375	.3035	.5589	.0000	.5833	.4167
trad. meat =	(T=)	.1900	.2606	.5494	.3657	.3685	.2658
trad. meat −	(T−)	.2080	.2192	.5728	.5823	.1366	.2811
naturalness +	(N+)	.2028	.3131	.4841	.2398	.6892	.0710
naturalness =	(N=)	.2001	.2489	.5510	.4484	.3063	.2453
naturalness −	(N−)	.1578	.2169	.6253	.3804	.0852	.5344
selectivity +	(Se+)	.2246	.2595	.5159	.5140	.3878	.0982
selectivity =	(Se=)	.1713	.2351	.5937	.3750	.2028	.4222
selectivity −	(Se−)	.1686	.2794	.5520	.2123	.4636	.3242
scepticism +	(Sk+)	.1975	.2524	.5502	.4252	.3251	.2498
scepticism =	(Sk=)	.1989	.2688	.5323	.3748	.4236	.2016
scepticism −	(Sk−)	.1421	.2257	.6321	.2851	.1256	.5894
luxury +	(L+)	.2186	.2068	.5745	.6692	.0710	.2597
luxury =	(L=)	.1586	.2647	.5768	.2206	.3688	.4106
luxury −	(L−)	.1922	.2907	.5171	.2721	.5483	.1796
versatility +	(V+)	.1780	.2159	.6061	.4683	.0946	.4372
versatility =	(V=)	.2044	.2857	.5100	.3404	.5276	.1320
versatility −	(V−)	.1564	.2538	.5898	.2485	.3029	.4486
latent budgets							
Pork		.7439	.0000	.2902	.3637	.2697	.2573
Beef		.0000	.0000	.5266	.2799	.2333	.3797
Horsemeat, Veal, Mutton		.0307	.0215	.0279	.0275	.0261	.0267
Mince		.0000	.4604	.1553	.1686	.2379	.2106
Processed meat		.0478	.4798	.0000	.1031	.1853	.1059
Other		.1776	.0383	.0000	.0572	.0478	.0197
budget proportions (π_k)		.1838	.2540	.5622	.3860	.3127	.3013

Now that we have identified the LBM we can offer an interpretation. The outer extreme solution can be found in the left panel of Table 5. The appropriate way to label the budgets is by their reference to the categories of the response variable.

$\hat{\beta}_1^*$ is dominated by pork and snacks/barbecue meat, while beef and mince are absent. Since most snacks in Holland are made from pork as well, the interpretation of this budget is rather straightforward and the budget can be described as a pork budget. $\hat{\beta}_2^*$ is characterized by mince and processed meat while the traditional kinds of meat pork and beef are absent. The characteristics of mince and most kinds of processed meat are the high fat level and their easy and fast preparation. The budget can be described in these terms. $\hat{\beta}_3^*$ is dominated by beef, pork and mince in approximately the quantities served in a traditional Dutch household. This budget can thus be described as a traditional budget.

The inner extreme solution can be found in the right panel of Table 5. The appropriate way to label the budgets is by their reference to participant groups. $\hat{\beta}_1^*$ is labeled the *conscious budget* and dominated by respondents who are conscious of both costs and their personal health, who do not like cooking, do not consider meat to be very important, but rather regard meat as a luxury product. The characteristics of $\hat{\beta}_2^*$ are the importance of ready-made products and their naturalness, a positive attitude towards traditional meat and cooking. Although the recognition of the importance of ready-made products contradicts the other characteristics the best way to summarize the attitudes typical for this budget is to label it a *traditional budget*. $\hat{\beta}_3^*$ is described as the *quality budget*. This is shown by its dominant characteristics, a positive attitude towards the importance and quality of meat, plus concerns about naturalness. The attitude towards health, however, is rather ambiguous.

4. Relation to Correspondence Analysis

A related model for the analysis of compositional data is correspondence analysis (CA; see for example Nishisato 1980, 1994; Greenacre 1984; Gifi 1990). We define CA(M + 1) as

$$\tilde{\pi}_{ij} = p_i p_j \left[1 + \sum_{m=1}^{M} \lambda_m r_{im} c_{jm} \right]$$

where the scores r_{im} and c_{jm} are restricted by $\Sigma_i p_i r_{im} = \Sigma_j p_j c_{jm} = 0$ and $\Sigma_i p_i r_{im}^2 = \Sigma_j p_j c_{jm}^2 = 1$. The elements $\tilde{\pi}_{ij}$ provide rank (M + 1) approximations of observed proportions p_{ij} $(0 \le M \le \min(I-1, J-1))$. $\tilde{\pi}_{ij}$ are collected in an $I \times J$ two-way matrix, $\tilde{\Pi}$. If $0 \le M < \min(I-1, J-1)$, then $\tilde{\Pi}$ has a reduced rank. The parameters λ_m are the singular values obtained from a singular value decomposition of the matrix with elements $(\tilde{\pi}_{ij} - p_i p_j)/(p_i p_j)^{1/2}$. Because of the mathematical properties of the singular value decomposition, the matrix $\tilde{\Pi}$ is an optimal approximation of the observed matrix in a (weighted) least-squares sense (see, for example, Greenacre 1984).

De Leeuw and van der Heijden (1991) show that a compositional data matrix of rank K can always be decomposed by CA(K), but not always by LBM(K) if $2 < k < \min(I-1, J-1)$. The geometric study in Section 2 demonstrated that a probability matrix of rank K can be decomposed if the budgets are a convex combination of K latent budgets. For example, a compositional data matrix can be decomposed by LBM(3) if and only if the budgets lie within a triangle in U_2 (cf. Figure 4). Therefore, for a compositional data matrix of rank $K = 3$, CA(K) and LBA(K) are equivalent if the budgets lie within a triangle.

The inadequacy of the estimation procedure proposed for LBA by Renner (1993) and criticized by van der Heijden (1994), can be illustrated in this way. Renner begins his estimation procedure for LBM(K) by determining a lower rank matrix, performing a singular value decomposition of the matrix of observed compositions. Assume that $K = 3$. His initial estimates need not fall into U_2, and the rows in the lower rank matrix need not be budgets since elements can be negative. Renner uses ad hoc procedures to adjust this, but the resulting matrix is no longer optimal in the least-squares sense. A singular value decomposition does not constrain the projections for the rows to fall within a triangle. The least-squares estimation procedure of Mooijaart and van der Heijden (1995) circumvents these problems by estimating the parameters of the LBM directly.

5. Discussion

In this paper, we have studied the geometry of the LBM. We found that in the model with K latent budgets, a valid solution is found if the $K-1$ dimensional polyhedron V^*_{K-1} spanned by the identified latent budgets is a subset of U_{K-1}, a $K-1$ dimensional polygon, spanned by the outer extreme budgets, and if V^*_{K-1} includes W_{K-1}, also a $K-1$ dimensional polygon spanned by the inner extreme budgets. A special case is LBM(2) where W_1, V^*_1 and U_1 all have the same shape, namely a one-dimensional line segment. In this case identification is straightforward. We proposed a general method to identify LBM's. by maximizing/minimizing the sum of chi-squared distances among the latent budgets. The identification method proposed earlier by de Leeuw *et al.* (1990) is a special case of this method. We believe that this criterion has desirable properties, such as easy interpretation of the latent budgets in the outer extreme solution, and maximal differences among the categories of the explanatory variable in the inner extreme solution. The proposed solutions are mathematical and it is likely that a theory-driven solution is easier to interpret. However, LBA is mainly used as an exploratorive technique and our identifiability may facilitate interpretation in the situation where we have no prior knowledge about the structure of the data.

References

AITCHISON, J. (1986), *The Statistical Analysis of Compositional Data*, London: Chapman and Hall.

BANDEEN-ROCHE, K. (1994), ''Resolution of Additive Mixtures into Source Components and Contributions: A Compositional Approach,'' *Journal of the American Statistical Association, 89*, 1450-1458.

CLOGG, C. C. (1981), ''Latent Structure Models of Mobility,'' *American Journal of*

Sociology, 86, 836-868.

DE LEEUW, J., and VAN DER HEIJDEN, P. G. M. (1987), "The Analysis of Time-Budgets with a Latent Time-Budget Model," in *Data Analysis and Informatics V.*, Ed., E. Diday, Amsterdam: North-Holland, 159-166.

DE LEEUW, J., and VAN DER HEIJDEN, P. G. M. (1991), "Reduced Rank Models for Contingency Tables," *Biometrika, 78*, 51, 229-232.

DE LEEUW, J., VAN DER HEIJDEN, P. G. M., and VERBOON, P. (1990), "A Latent Time Budget Model," *Statistica Neerlandica, 44*, 1-21.

EVERITT, B. S., and HAND, D. J. (1981), *Finite Mixture Distributions*, London: Chapman and Hall.

GIFI, A. (1990), *Non-linear Multivariate Analysis*, New York: Academic Press.

GOODMAN, L. A. (1974a), "The Analysis of Systems of Qualitative Variables when Some of the Variables Are Unobservable. I. A Modified Latent Structure Approach," *American Journal of Sociology, 79*, 1179-1259.

GOODMAN, L. A. (1974b), "Explanatory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models," *Biometrika, 61*, 215-231.

GREENACRE, M. J. (1984), *Theory and Applications of Correspondence Analysis*, London: Academic Press.

LINDSAY, B. G. (1995), *Mixture Models: Theory, Geometry and Applications*, (NSF-CBMS Regional Conference Series in Probability and Statistics Volume 5), Hayward, CA: Institute of Mathematical Statistics.

METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H., and TELLER, E. (1953), "Equations of State Calculations by Fast Computing Machines," *Journal of Chemical Physics, 21*, 1087-1091.

MOOIJAART, A., and VAN DER HEIJDEN, P. G. M. (1995), *A Least Squares Algorithm for a Mixture Model for Compositional Data*, (Methods Series MS-95-2), Utrecht: Utrecht University, Faculty of Social Sciences, ISOR.

NISHISATO, S. (1980), *Analysis of Categorical Data: Dual Scaling and its Applications*, Toronto: Toronto University Press.

NISHISATO, S. (1994), *Elements of Dual Scaling: An Introduction to Practical Data Analysis*, Hillsdale, NJ: : Lawrence Erlbaum.

PRESS, H. P., FLANNERY, B. P., TEUKOLSKY, S. A., and VETTERLING, W. T. (1989), *Numerical Recipes in Pascal*, Cambridge: Cambridge University Press.

RENNER, R. M. (1993), "The Resolution of a Compositional Data Set into Mixtures of Fixed Source Compositions," *Applied Statistics, 42*, 615-631.

RENNER, R. M. (1995), "The Construction of Extreme Compositions," *Mathematical Geology, 74*, 485-497.

RIPLEY, B. D. (1990), *Unmixing Finite Mixtures*, Unpublished BP-research report, Sunbury.

TITTERINGTON, D. M., SMITH A. F. M., and MAKOV, U. E. (1985), *Statistical Analysis of Finite Mixture Distributions*, New York: Wiley.

VAN DER ARK, L. A., and VAN DER HEIJDEN, P. G. M. (1998), "Graphical Display of Latent Budget Analysis and Latent Class Analysis, with Special Reference to Correspondence Analysis," in *Visualization of Categorical Data*, Eds., J. Blasius, and M. Greenacre, San Diego: Academic Press, 489-509.

VAN DER HEIJDEN, P. G. M. (1994), "End-member Analysis and Latent Budget Analysis," *Applied Statistics, 43, 3*, 527-30.

VAN DER HEIJDEN, P. G. M., MOOIJAART, A., and DE LEEUW, J. (1992), "Constrained Latent Budget Analysis, in *Sociological Methodology 1992*, Ed., P. V. Marsden, Cambridge: Blackwell, 279-320.