Title
A Latent Time-Budget Model

Authors
Leeuw, Jan de
Heijden, Peter G. M. van der
Verboon, Peter

Publication Date
1990-05-03

Peer reviewed

A latent time-budget model

J. de Leeuw
Departments of Psychology and Mathematics
U.C.L.A., Hilgard Ave. 405
Los Angeles CA 90024-1563 USA

P.G.M. van der Heijden
F.S.W., Department V.O.S., Section Methods
University of Utrecht
Postbus 80140, 3508 TC Utrecht
The Netherlands
and

P. Verboon
Department of Data Theory University of Leiden
Wassenaarseweg 52, 2333 AK Leiden
The Netherlands

Time-budgets summarize how the time of objects is distributed over a number of categories. Usually they are collected in object by category matrices with the property that rows of this data matrix add up to one. In this paper we discuss a model for the analysis of time-budgets that used this property. The model approximates the observed time-budgets by weighted sums of a number of latent time-budgets. These latent time-budgets determine the behavior of all objects. Special attention is given to the identification of the model. The model is compared with logcontrast principal component analysis.

Key Words & Phrases: Latent structure, factor analysis, data analysis, EM-algorithm, identication, compositional data.

Writing this paper was partly made possible by a grant of the Netherlands Organisation for the Advancement of Pure Research (N.W.O.) for the second author.

1. Introduction

Time-budgets can be collected if one is interested in the way in which the time of objects (for example, persons, animals, countries) is distributed over a number of distinct, non-overlapping activities. This type of information can be collected in a two-way matrix, with objects in the rows and activities in the columns. In a cell we find the proportion of time that a specific object has spent on a specific activity. Each row of this matrix adds up to 1.

In table 1 we find an example of a matrix with time-budget data. The data are derived from GROSS et al. (1985) and were analyzed earlier by the DE LEEUW and VAN DER HEIJDEN (1988) and VAN DER HEIJDEN (1987). The objects are males, females and children of four Indian tribes from the Amazone. The tribes are the Mekranoti, the Kanela, the Bororo and the Xavente. Six activities were recorded between 6.00 A.M. and 8.00 P.M., namely 'being idle', 'sleeping', 'caring', 'nonsubsistence behavior', 'domestic activities' and the activity 'wild' (hunting, fishing, gathering). For more details on the definition of these categories we refer to GROSS et al. (1985). Unfortunately, we were not able to obtain the original data, that were collected with the random spot check method (see below). The percentages in the cells of the matrix in table 1 were derived from stacked bar histograms.

TABLE 1. Time-budgets of Amazone Indians.

Activities: 1 = being idle; 2 = sleeping; 3 = caring; 4 = non-subsistence behavior; 5 = domestic activities; 6 = 'wild'.

Activities:		1.	2.	3.	4.	5.	6.	Tot
Mekranoti	Males	.463	.056	.006	.278	.070	.128	1.000
	Females	.434	.068	.074	.206	.199	.019	1.000
	Children	.733	.079	.007	.029	.130	.021	1.000
Kanela	Males	.562	.037	.016	.201	.125	.060	1.000
	Females	.492	.046	.097	.269	.093	.004	1.000
	Children	.789	.107	.006	.032	.062	.004	1.000
Bororo	Males	.563	.049	.002	.142	.121	.124	1.000
	Females	.489	.046	.077	.321	.032	.034	1.000
	Children	.748	.109	.002	.071	.033	.037	1.000
Xavente	Males	.660	.036	.003	.127	.097	.078	1.000
	Females	.423	.023	.095	.335	.089	.035	1.000
	Children	.767	.068	.005	.080	.066	.014	1.000

A matrix with time-budget data can be derived in distinct ways. Firstly, time-budgets can be estimated with the so-called 'random spot check method' (see GROSS, 1984). In this method we know at random points in time what an object is doing. The number of times that we observe a specific activity for some object gives an indication for the total amount of time that this object spends on this activity. The idea is simply that when, for example, a person is baking bread 20 times out of the 100 times that we have observed him, we estimate that he is baking bread 20% of his time. The random spot check method stems from anthropology, but is also used in, for example, ethology, where the activities of the objects are denoted by some observer. In the human sciences objects can play a more active role, for example, persons can denote their current activity each time some beeper beeps (see, for example, ROBINSON, 1985).

Another method to get time-budget data is by deriving them from so-called event-history data. For event-history data we know for each object the length of time and the order of the activities. This type of data can be collected, for example, by asking persons to keep a diary. This type of data can also be collected in a matrix like in table 1. In this summary the time order of the respective activities is lost.

The analysis of time-budget data has the objective to study the relation between the objects on the one hand and the activities on the other. The following research questions could be important: are there objects with a clearly deviating pattern of activities? Is it possible to find groups of objects with very similar patterns of activities? Are there activities that only occur for a long period for some specific objects? Van der Heijden (1987) gives an overview of data analysis methods stemming from distinct disciplines such as ethology, sociology, anthropology to tackle such questions. In the first place such data can be displayed using histograms, see figure 1.

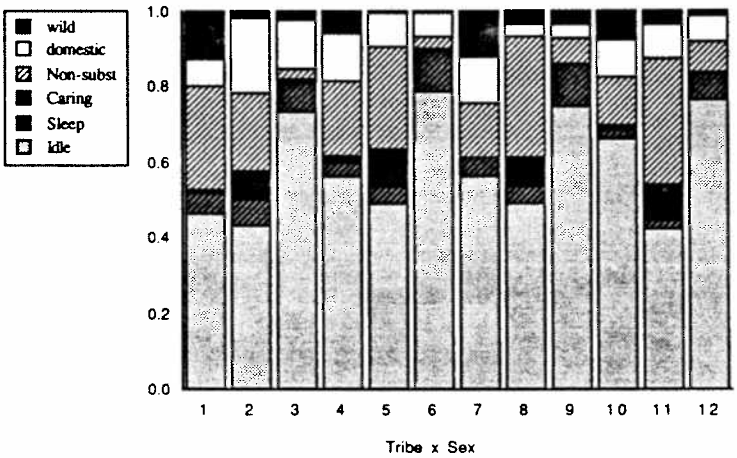

FIGURE 1. A stacked bar histogram of the data in table 1.

This gives a good representation of the 'better filled' activities, but on the whole shows little from the possible relations in the data matrix. Another drawback of this method is that, when the number of activities or objects becomes large, the histogram becomes difficult to use as a summary. E

In much research we can find tables and graphical representations, in all sorts of ways (see, for example, SZALAI, 1972; PARKES and THRIFT, 1980; STAIKOV, 1982; HARVEY et al., 1984). Limited as these methods may be, a clear advantage of this approach is that the results are easily explained to a large audience. Some researchers use standard multivariate methods, where correlations between the columns of the two-way matrix are calculated (see ELLIOTT, 1984). The drawback of this is that correlations are not very well suited as measures for the similarity of columns of frequencies, and no use is made of the special property of time-budget data that each row of the matrix adds up to 1 (see AITCHISON, 1986, for an extensive discussion of these points).

Another possibility is to use methods that are specifically devised for the analysis of contingency tables, such as correspondence analysis (GREENACRE, 1984). Here both objects and activities are displayed graphically as points in a multi-dimensional space, where distances between points indicate in which way the matrix departs from independence. In the context of time-budget data attention for correspondence analysis is growing (see, for example, SAPORTA, 1981, 1985; DE LEEUW et al., 1985; VAN DER HEIJDEN, 1987; VAN DER HEIJDEN and DE LEEUW, 1989). Correspondence analysis seems to be very well suited for this type of data with rows that add up to 1, because the distance measure that is used by correspondence analysis, the so-called chi-squared distance, indicates how rows (columns) with conditional proportions deviate from each other. A drawback is, however, that the inferential side of correspondence analysis is not very well developed (unless one uses models recently proposed by GOODMAN, 1985, 1986) and that the results are a bit difficult to handle by laymen (cf. DE LEEUW and VAN DE HEIJDEN, 1988). Loglinear analysis methods have the drawback that, if there is a relation between the objects and the activities, the number of parameters to be interpreted increases rapidly, and that the parameters do not reflect the fact that the proportions in each row add up to 1.

An important step forward in the analysis of time budget data was made by AITCHISON (1986). He studied the analysis of a broader class of data, namely compositional data, of which time-budget data are a special case. Compositional data can be collected in a data matrix in which rows are vectors of proportions that each add up to one. He considers many types of compositional data, and many types of questions that can be asked about these data. Basically, his approach is to transform these data so that it is possible to work in R instead of R^+, and to concentrate on the covariances between the columns under the assumption that the rows of the transformed matrix are replications from a multinormal distribution. In the discussion of this paper we will compare the method that we propose for the analysis of time budgets with his method for dimension reduction of compositional data coined "logcontrast principal component analysis" (AITCHISON, 1986, chapter 8).

In this paper we discuss another model for the analysis of time-budgets that has the advantage that it shows in a simple way what relations exist between the objects and the activities (see also DE LEEUW and VAN DER HEIJDEN, 1988). The model uses the special properties of time-budget data, and, if the data are collected under product-multinomial sampling (for example, with the random spot check method), it is possible to evaluate the fit of the model.

2. THE MODEL

We start from an $I \times J$ matrix with objects $i (i = 1,...,I)$ in the rows and activities $j (j = 1,...,J)$ in the columns. We assume that the data are collected with the random spot check method. The number of times that object i was doing activity j is denoted by n_{ij}. The total number of observations for object i is fixed by the design of the study and equal to n_{i+} (an index is replaced by '+'

if we add up over the corresponding way of the matrix). We can derive observed conditional proportions $p_{ij} = n_{ij}/n_{i+}$, and the observed conditional proportions p_{ij} can be considered as estimates of π_{ij}, the theoretical conditional proportion of time spent by object i on activity j in a typical period for object i. If object i represents a group, then π_{ij} can be interpreted as the theoretical proportion of time on activity j for a typical group member.

If we assume that the observations are independent, then the n_{ij} in row i follow a multinomial distribution with $E(n_{ij}) = n_{i+}\pi_{ij}$. This assumption is only realistic if the intervals between the observations are large enough. If the objects do not influence each other, we can assume that the n_{ij} in the matrix follow a product-multinomial distribution. In the sequel we will assume that this assumption is not too heavily violated. Now we have the possibility to test whether the distributions of conditional proportions p_{ij} differ. With this aim we calculate expected proportions under the model of independence. Assuming that the proportions in table 1 are based on 100 observations per row, the likelihood ratio statistic G^2 equals 251.9 with $(I-1)(J-1)=55$ degrees of freedom. The statistic is asymptotically chi-squared distributed under the nulhypothesis. We conclude that it is significant, and therefore we have to reject the model. The various rows of observed conditional proportions do not stem from one underlying row of theoretical conditional proportions.

In interesting applications independence of objects and activities will be rejected almost always. Therefore we propose to study the form of the departure from independence with a model for latent time-budgets. This model for the theoretical conditional proportions π_{ij} (where $\pi_{i+} = 1$) has the following form:

$$\pi_{ij} = \sum_{k=1}^{K} \alpha_{ik}\beta_{jk} \tag{1}$$

Here $k(k = 1,...,K)$ indexes the K latent budgets, where K is fixed by the researcher. The idea behind this model is that there are K typical or latent time-budgets specified by $(\beta_{1k},...,\beta_{Jk})$ that determine the behavior of all objects. These parameters for the activities add up to one for each latent budget and can be interpreted as proportions, i.e. $0 \leqslant \beta_{jk} \leqslant 1$ and $\beta_{+k} = 1$. Each latent time-budget k 'explains' the time spending behavior of each object i to some extent, specified by α_{ik}. These parameters add up over the latent budgets to one for each object, and can be interpreted as proportions, i.e. $0 \leqslant \alpha_{ik} \leqslant 1$ and $\alpha_{i+} = 1$. Notice that for $K = 1$ this model reduces to the independence model.

The number of degrees of freedom is $(I-K)(J-K)$. This formula for the number of degrees of freedom is derived below. In DE LEEUW and VAN DER HEIJDEN (1988) and van der HEIJDEN (1987) this number was given incorrectly.

ESTIMATION

In this section we derive an algorithm to compute ML estimates of the parameters of our model. It is closely related to the EM-algorithm (DEMPSTER, LAIRD and RUDIN, 1977), but we give a derivation which uses the idea of majorization taken from convex analysis. Since we assume that the observations are collected under a product-multinomial distribution the likelihood function L is given by

$$L = \sum_{i=1}^{I} n_{i+} \sum_{j=1}^{J} p_{ij} \ln \pi_{ij} = \sum_{i=1}^{I} n_{i+} \sum_{j=1}^{J} p_{ij} \ln \sum_{k=1}^{K} \alpha_{ik}\beta_{jk}$$

Now suppose $\underline{\alpha}_{ik}$ and β_{jk} are the current values of the parameters, giving estimated values $\underline{\pi}_{ij}$. By concavity of the logarithm

$$\ln \pi_{ij} - \ln \underline{\pi}_{ij} \geqslant \{ \sum_{k=1}^{K} \underline{\alpha}_{ik}\underline{\beta}_{jk} \ln \alpha_{ik}\beta_{jk} - \sum_{k=1}^{K} \underline{\alpha}_{ik}\underline{\beta}_{jk} \ln \underline{\alpha}_{ik}\underline{\beta}_{jk} \} / \underline{\pi}_{ij}.$$

For convenience we define $\pi_{ijk} = \underline{\alpha}_{ik}\beta_{jk}$ and $\pi_{ijk} = \alpha_{ik}\ \beta_{jk}$. Also $\underline{L}$ is the value of the likelihood function at the current parameters. Then

$$L \geqslant \underline{L} + \sum_{i=1}^{I} n_{i+} \sum_{j=1}^{J} p_{ij} \{ \sum_{k=1}^{K} \underline{\alpha}_{ik}\underline{\beta}_{jk} \ln \alpha_{ik}\beta_{jk} - \sum_{k=1}^{K} \underline{\alpha}_{ik}\underline{\beta}_{jk} \ln \underline{\alpha}_{ik}\underline{\beta}_{jk} \} / \underline{\pi}_{ij} \tag{2}$$

with equality if and only if $\underline{\pi}_{ijk} = \pi_{ijk}$ for all i,j,k. Now suppose we maximize the right hand side of (2) over α_{ik} and β_{jk} (with the normalization restrictions). Then L at the maximum is larger than the maximum of the right hand side, which is larger than the value of the right hand side at $\underline{\alpha}_{ik}$ and β_{jk} which is L. Thus maximizing the right hand side of (2) actually increases the likelihood, and is really simple, because it amounts to maximizing

$$\mathbf{L} = \sum_{i=1}^{I}\sum_{k=1}^{K} n_{i+k} \ln \alpha_{ik} + \sum_{j=1}^{J}\sum_{k=1}^{K} n_{+jk} \ln \beta_{jk} \tag{3}$$

with $n_{ijk} = n_{i+} p_{ij} \alpha_{ik}\beta_{jk} / \underline{\pi}_{ij}$. Note that $n_{ij+} = n_{i+} p_{ij} = n_{ij}$.

Now we find the estimates of expected frequencies as follows. Consider the observed frequencies $n_{ij} = n_{i+} p_{ij}$. We start with trial values for $\underline{\alpha}_{ik}$ and $\underline{\beta}_{jk}$. Then calculate

$$n_{ijk} = n_{i+} p_{ij} \underline{\alpha}_{ik}\underline{\beta}_{jk} / \underline{\pi}_{ij}, \tag{4}$$

and after this update the parameter estimates by

$$\alpha^{\bullet}_{ik} = n_{i+k} / n_{i++} \tag{5}$$

$$\beta^{\bullet}_{jk} = n_{+jk} / n_{++k} \tag{6}$$

which ends the first cycle of the algorithm. In each cycle the current best estimates of the expected frequencies $m_{ij} = n_{i+} \pi_{ij}$ can be used to derive the value of the likelihood ratio statistic G^2. When the difference between two subsequent G^2-values is smaller than a prespecified criterion, iterating can stop, and we can test the fit of the model using the current value of G^2.

Two types of parameter constraints are easily built into the algorithm, namely *specific value constraints* and *equality constraints*. It amounts to maximizing (3) under the constraints imposed. First we consider specific value restrictions, for example, $\alpha_{ik}=c$ or $\beta_{jk}=c$ for some i and k, or j and k, where $0 \leqslant c \leqslant 1$. This restriction should be imposed in each iteration after step (6) of the algorithm. Equality constraints can be built in in a similar way. For β_{jk}-parameters we have to use the weighted average of two (or more) $(n_{++k}/n_{+++})\beta_{jk}$ that are constrained to be equal. For α_{ik}-parameters we should use the weighted averages of two $(n_{i+}/n_{++})\alpha_{ik}$ that are constrained to be equal. Specific value constraints and equality constraints can also be built in simultaneously. After imposing constraints, it is necessary to adjust the other parameters in each iteration in order to satisfy $\alpha_{i+}=1$ and $\beta_{+k}=1$.

The algorithm can converge to a non-global maximum. In order to be confident that one has reached a global maximum and not a local maximum, different starting values should be used.

The estimation procedure shows the following properties of the model:

PROPERTY 1. At a stationary point the proportion of observations falling into latent budget k is $\pi_k \equiv n_{++k}/n_{+++}$. Hence proportion π_k is a measure of the *importance* of latent budget k.

PROPERTY 2. The decomposition $\pi_{ij}=\Sigma_k \alpha_{ik}\beta_{jk}$ with restrictions $\alpha_{i+}=1$, $\beta_{+k}=1$ and $\pi_{i+}=1$ is equivalent to the decomposition of the transposed matrix with elements $\pi^*_{ij}=\Sigma_k \alpha^*_{ik}\beta^*_{jk}$ with restrictions $\alpha^*_{+k}=1$, $\beta^*_{j+}=1$ and $\pi^*_{+j}=1$.

PROOF. $n_{i+}\pi_{ij}=n_{i+}\Sigma_k \alpha_{ik}\beta_{jk}= n_{i+}\Sigma_k(n_{i+k}/n_{i++})(n_{+jk}/n_{++k})= \Sigma_k (n_{i+k}n_{+jk}/n_{++k})$. Similarly, $n_{+j}\pi^*_i= n_{+j}\Sigma_k \alpha^*_{ik}\beta^*_{jk}= n_{+j}\Sigma_k(n_{i+k}/n_{++k})(n_{+jk}/n_{+j+})= \Sigma_k(n_{i+k}n_{+jk}/n_{++k})$. Both decompositions provide us with the same expected frequencies: $m_{ij}=n_{i+}\pi_{ij}=n_{+j}\pi^*_i$. The decomposition of π^*_{ij} is derived from that of π_{ij} by $\beta^*_{jk}=\pi_k\beta_{jk}/\Sigma_k\pi_k\beta_{jk}$ and $\alpha^*_{ik}=n_{i+}\alpha_{ik}/\Sigma_i n_{i+}\alpha_{ik}$.

PROPERTY 3. At a stationary point $\Sigma_i n_{i+}\pi_{ij}=n_{+j}$.

PROOF. This is easy to see using that $n_{ij+}=n_{ij}$: $\Sigma_i n_{i+}\pi_{ij}=\Sigma_i\Sigma_k n_{i+}\alpha_{ik}\beta_{jk}= \Sigma_k\Sigma_i n_{i+}(n_{i+k}/n_{i++})(n_{+jk}/n_{++k})= \Sigma_k(\Sigma_i n_{i+k})(n_{+jk}/n_{++k})=n_{+j}$.

PROPERTY 4. At a stationary point $\Sigma_k\pi_k\beta_{jk}=n_{+j}/n_{++} \equiv \pi_j$.

PROOF. $\Sigma_k\pi_k\beta_{jk}= \Sigma_k(n_{++k}/n_{+++})(n_{+jk}/n_{++k})=n_{+j+}/n_{+++}=n_{+j}/n_{++}$.

EXAMPLE. We estimated the model with $K=2$ latent time-budgets for the example in table 1. The estimates for the parameters are in table 2. When we assume that proportions in each row are derived from 100 observations, the fit is $G^2=$ 96.3 with 40 degrees of freedom, which is significant. The model with $K=3$ latent time-budgets gives a much better fit, namely $G^2=$ 37.0 for df

$=27$, which is not significant at $p=.05$. We conclude that the model gives an adequate description of the data. The parameters for $K=3$ can also be found in table 2.

TABLE 2. Parameters estimates for K = 1,2,3.

Row parameters:		K=1	K=2		K=3		
		k=1	k=1	k=2	k=1	k=2	k=3
Mekranoti	Males	1.000	.239	.761	.002	.947	.051
	Females	1.000	.277	.723	.357	.110	.533
	Children	1.000	.946	.054	.865	.114	.021
Kanela	Males	1.000	.443	.557	.319	.512	.169
	Females	1.000	.220	.780	.360	.003	.637
	Children	1.000	.962	.038	.932	.029	.039
Bororo	Males	1.000	.543	.457	.275	.719	.006
	Females	1.000	.174	.826	.239	.218	.543
	Children	1.000	.893	.107	.753	.245	.003
Xavente	Males	1.000	.670	.330	.456	.531	.013
	Females	1.000	.020	.980	.120	.221	.659
	Children	1.000	.866	.134	.795	.140	.064
Activities							
	Being idle	.594	.781	.390	.817	.484	.299
	Sleeping	.060	.087	.031	.095	.040	.021
	Caring	.032	.000	.068	.002	.000	.138
	Nonsubsistence	.174	.023	.339	.006	.238	.421
	Domestic	.093	.081	.105	.080	.094	.116
	'Wild'	.047	.028	.067	.000	.143	.005

We can describe the most important results as follows. A simple way to interpret the parameters is to start with the interpretation of the latent budgets, by comparing the latent budget elements β_{jk} with the marginal proportions π_j (i.e. the latent budget $K=1$). We do this for the solution with $K=3$ latent budgets only. The first latent budget is typical for an activity pattern in which persons are much more than average idle (.817 as compared with .594) and sleeping (.095 versus .060). The other activities occur much less than average. The second latent budget is typical for a behavior pattern in which persons perform much more than average nonsubsistence activities (.238 versus .174) and the activity 'wild'. (.143 versus .047). Being idle, sleeping and caring are performed less often than average. The third latent budget is typical for caring (.138 versus .032) and nonsubsistence activities (.421 versus .174). Being idle, sleeping and 'wild' are performed much less often than average. Now that we know how to interpret the latent budget, we can interpret how the behavior of the groups of persons is built up from them. A first idea can be obtained by looking for the highest proportion in each row of contributions α_{ik}. Roughly, we see that the behavior of the children of each tribe is built up for the largest part from the first latent budget, i.e. they are much more than average idle and sleeping. The males load highest on this second latent budget, showing that they perform much more than average nonsubsistence activities and the category 'wild'. The pattern of activities of the females is built up for the largest part from the third latent budget, i.e. from caring and nonsubsistence activities. We conclude that the differences between the males, females and children dominate the solution, and seem to be larger than the differences between tribes. However, we can also discern tribe differences, for example, the activity pattern of the Mekranoti males is clearly dominated by the second latent budget (.947), whereas this proportion is much lower for the Kanela males and the Xavente males.

If we plot each object i in a three-dimensional space with its parameters α_{ik} as coordinates, then the cloud of object points falls into a triangular two-dimensional subspace within corner points (1,0,0), (0,1,0) and (0,0,1). See figure 2. It shows in a clear way what is going on for the row parameters in table 2 for $K=3$. With each point in figure 2 a budget is associated (see table 3): each of the 12 objects i has an *expected* budget $(\pi_{i1},...,\pi_{iJ})$ associated with it, the *average* has $\pi_1,...,\pi_j,...,\pi_J$ as its budget, and each corner point k has its *latent* budget $(\beta_{1k},...,\beta_{Jk})$. The expected budget for row i can be derived in a simple way from the three latent budgets, for example, the budgets for rows 7, 9 and 10 are somewhere in between the latent budgets 1 and 2. The average budget can be derived from the data directly as $\pi_j = \Sigma_i(n_{i+}/n_{++})\pi_{ij}$. They can also be derived from the latent budgets using property 4. The proportions π_k (see Property 1) are used as coordinates for the average budget in figure 2.

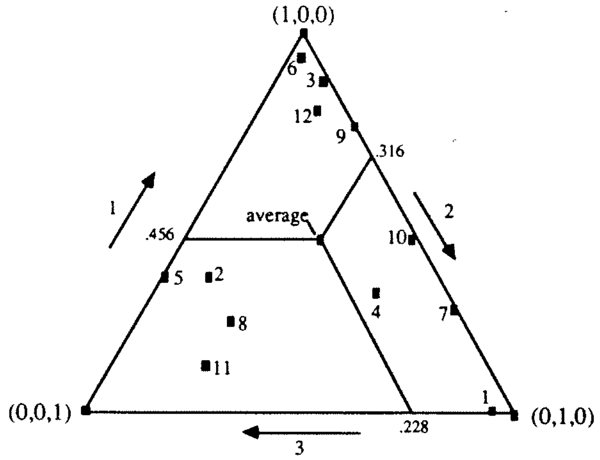

FIGURE 2. Plot of the α-parameters for K=3. Example: coordinates of average are .456, .316, .228.

TABLE 3. Expected time-budgets for K=3.

Activity numbers as in table 1.

Activities:		1.	2.	3.	4.	5.	6.	Tot
Mekranoti	Males	.475	.039	.007	.247	.095	.136	1.000
	Females	.504	.049	.074	.253	.101	.019	1.000
	Children	.768	.087	.005	.041	.083	.016	1.000
Kanela	Males	.559	.054	.024	.195	.094	.074	1.000
	Females	.486	.047	.089	.271	.103	.004	1.000
	Children	.787	.090	.007	.029	.082	.004	1.000
Bororo	Males	.574	.055	.001	.176	.091	.103	1.000
	Females	.463	.042	.075	.282	.103	.034	1.000
	Children	.734	.081	.002	.064	.084	.035	1.000
Xavente	Males	.633	.064	.003	.135	.088	.076	1.000
	Females	.402	.034	.091	.331	.107	.035	1.000
	Children	.737	.082	.011	.065	.085	.020	1.000
Latent budget 1		.817	.095	.002	.006	.080	.000	1.000
Latent budget 2		.484	.040	.000	.238	.094	.143	1.000
Latent budget 3		.299	.021	.138	.421	.116	.005	1.000
Average budget		.594	.060	.032	.174	.093	.047	1.000

We can make a similar plot for the column parameters β_{jk} when we use β^{*}_{jk} derived in Property 2. See figure 3. We can deduce from figure 3 that activities 1 (idle) and 2 (sleep) fall more than average in the first latent budget; 6 (wild) is more than average in the second latent budget 6; 3 (caring) and 4 (non-subsistence activities) are more than average in the third latent budget.

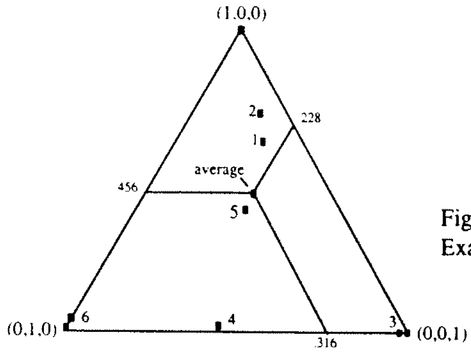

Figure 3. Plot of the rescaled β–parameters for K=3. Example: coordinates of average are .456, .316, .228

The goal of the model should be clear by now: it tries to give a parsimonious description of the data with typical time-budgets given by β_{jk}. The proportions α_{ik} show how the objects load on these budgets, or, to put it differently, how much each typical budget explains from the time spending behavior of each object. The latent budgets describe the 'extreme' typologies between which each object falls.

3. IDENTIFICATION

The latent budget model is not identified. This aspect was not discussed by DE LEEUW and VAN DER HEIJDEN (1988) and VAN DER HEIJDEN (1987). One way to study identification is to construct the matrix with partial derivatives of the probabilities with respect to the parameters, and to determine if it is of full column rank. We will consider another, and in most respects more satisfactory, way to study the identification of the model. Consider (1). In matrix notation we can write (1) as $\Pi = AB'$. Model (1) is comparable with the reduced rank or singular value decomposition of matrices, used in factor analysis or principal component analysis. The main difference, or course, are the nonnegativity constraints on A and B. In reduced rank decomposition there are identification problems. A similar situation holds for the latent time-budget model: given matrices A and B, we can find new matrices $\tilde{A} = AT$ and $\tilde{B} = B(T^{-1})'$ that give an equivalent solution.

It is clear that, *if* (A,B) and $(\tilde{A},\tilde{B})$ satisfy the equations above, *then* (A,B) and $(\tilde{A},\tilde{B})$ provide equivalent solutions in the sense that $\Pi = AB' = \tilde{A}\tilde{B}'$. The reverse is proven as follows. Let Π have rank K. Then the columns of Π are J vectors of in R_I^+, and in this space they span a subspace with dimensionality K. The K columns of A and tue K columns of $\tilde{A}$ each span a cone in this same nonnegative subspace of dimensionality K. All columns of Π are in the cone spanned by A and in the cone spanned by $\tilde{A}$. Since both B and $\tilde{B}$ are nonnegative, the columns of Π can be derived both from A and $\tilde{A}$ as nonnegative linear combinations. Hence the columns of $\tilde{A}$ can be derived as linear combinations from the columns of A, i.e. $\tilde{A} = AT$. So, *if* (A,B) and $(\tilde{A},\tilde{B})$ are equivalent solutions, *then* the matrices are related by $\tilde{A} = AT$ and $\tilde{B} = B(T^{-1})'$.

So if we want to study the identification problem, we can do this by studying the matrix T. Starting from A and B we can derived new matrices $\tilde{A}$ and $\tilde{B}$. The new matrices $\tilde{A}$ and $\tilde{B}$ should have restrictions similar to

$$\sum_k \alpha_{ik} = 1 \tag{7a}$$

$$\sum_j \beta_{jk} = 1 \tag{7b}$$

$$0 \leqslant \alpha_{ik} \leqslant 1 \tag{7c}$$

$$0 \leqslant \beta_{jk} \leqslant 1 \tag{7d}$$

Therefore we cannot choose an arbitrary matrix T.

THEOREM 1. *Let u_I be a unit column vector of order I, let u_J have order J and let u_K have order K. Suppose A and B satisfy $Au_k = u_I$ and $B'u_J = u_K$, and A is of rank K, then $\tilde{A} = AT$ and $\tilde{B} = B(T^{-1})'$ satisfy $\tilde{A}u_k = u_I$ and $\tilde{B}'u_J = u_K$ if and only if $Tu_K = u_K$.*

PROOF.
1. If $Tu_K = u_K$, then $\tilde{A}u_K = ATu_K = u_I$.
2. If $\tilde{A}u_K = u_I$, then $ATu_K = u_I$. But also $Au_K = u_I$. Therefore $A(Tu_K - u_K) = 0$ and $Tu_K = u_K$ because A is of rank K.
3. If $Tu_K = u_K$ then $\tilde{B}'u_J = u_K$. This follows from $\tilde{B}'u_J = T^{-1}B'u_J = T^{-1}u_K$. We know that $Tu_K = u_K$ and therefore $T^{-1}u_K = u_K$.
4. If $\tilde{B}'u_J = u_K$ then $T^{-1}B'u_J = u_K$. Thus $T^{-1}u_K = u_K$ and therefore $u_K = Tu_K$. QED

We conclude that if we choose T so that T adds up to 1 rowwise, then $\tilde{A}$ will add up to 1 rowwise, and $\tilde{B}$ will add up to 1 columnwise, thus satisfying (7a) and (7b).

What further assumptions have to be imposed upon T so that the new $\tilde{A}$ and $\tilde{B}$ fulfil assumptions (7c) and (7d)? We will show this for our example, with $K = 2$ latent budgets. We only have to specify ranges for two values of T, since T adds up to one rowwise: so T has elements

$$T = \begin{bmatrix} x & 1-x \\ y & 1-y \end{bmatrix}.$$

We will first consider the ranges of x and y that are allowed to let the first column of $\tilde{A}$ satisfy (7c) (then the second column also satisfies (7c)). So, using table 2, we find

$$0 \leqslant .239x + .761y \leqslant 1 \tag{8a}$$

....

....

$$0 \leqslant ..866x + ..134y \leqslant 1 \tag{8l}$$

We now consider the equations for the restrictions on the values of $\tilde{B}$ that have to be fulfilled so that restriction (7d) holds. T^{-1} is given by

$$T^{-1} = \begin{bmatrix} (y-1)/(y-x) & (1-x)/(y-x) \\ (y)/(y-x) & (-x)/(y-x) \end{bmatrix}.$$

So, using the values of B displayed in table 2a we can find the following equations:

$$0 \leqslant .781(y-1)/(y-x) + .390(1-x)/(y-x) \leqslant 1 \tag{9a}$$

....

....

$$0 \leqslant .028(y-1)/(y-x) + .067(1-x)/(y-x) \leqslant 1 \tag{9f}$$

$$0 \leqslant .781(y)/(y-x) + .390(-x)/(y-x) \leqslant 1 \tag{10a}$$

....

....

$$0 \leqslant .028(y)/(y-x) + .067(-x)/(y-x) \leqslant 1 \tag{10f}$$

Many of the above inequalities are redundant, because, if we assume only that all values are larger than 0, then no value can be larger than 1 due to restrictions (7a) and (7b). If we work out all the inequalities, we find permissible (x,y) values that we can represent in a two-dimensional plane, see figure 4.

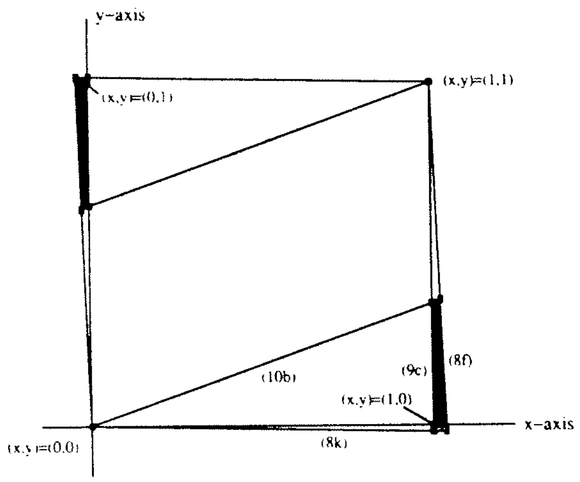

FIGURE 4. Ranges of permissible (x,y) values due to restrictions (7c) and (7d) in two shaded areas.

It shows two blocked areas of the same form with permissible values for (x,y). The two areas have the same shape because, by choosing an appropriate $\tilde{T}$, it is possible to interchange the first and the second column of $\tilde{A}$ and $\tilde{B}$. The values $(x,y)=(1,0)$ and $(x,y)=(0,1)$ are necessarily included in the areas, since they specify the identity matrix (i.e. we find $A=\tilde{A}$), and the situation where the columns of A interchange. Without loss of generality we concentrate on the lower right area. E

The lower right area is defined by four lines. Each of these lines corresponds with a specific parameter being equal to zero. We will sketch a fast procedure to find these lines.

We start with the two lines crossing right below the point $(x,y)=(1,0)$. These lines correspond with two α_{ik}-parameters being zero. These lines can be found by using only the extreme rows of A, i.e. the row whose values $(\alpha_{i1},\alpha_{i2})$ are nearest (0,1) and the row whose values $(\alpha_{i1},\alpha_{i2})$ are nearest (1,0). For our example this corresponds with row 11 (.020,.098) and row 6 (.962,.038), the other rows being inbetween. The reason that we only need two rows of A is that all inequalities should hold simultaneously, and two inequalities are most extreme. There are always two extreme inequalities, that define the lines that cross right below the point $(x,y)=(1,0)$. Say the line through $(x,y)=(1,1)$ has slope z, the line through $(x,y)=(0,0)$ has slope z^*. We are looking for the *smallest* z and for the *largest* z^*. The slope for row i is $-\alpha_{i1}/\alpha_{i2}$. It is easily seen that $-\infty \leqslant z < z^* \leqslant 0$. It follows that equations (8a) to (8l) restrict the area with permissible (x,y)-values by two lines that cross somewhere bottom right from $(x,y)=(1,0)$.

The other two lines, that cross top left from $(x,y)=(1,0)$, correspond with two β_{jk} being zero. These lines can be found as follows. For the line through $(x,y)=(1,1)$ we search for the largest slope z^{**} derived from the inequalities specified by equations (9a) to (9f); and for the line through $(x,y)=(0,0)$ we search for the smallest slope z^{***} derived from the inequalities specified by equations (10a) to (10f). The slope for category j is β_{j2}/β_{jl}. Hence the largest slope z^{**} is given by category 3, being $z^{**}=\infty$. The smallest slope z^{***} is given by category 2, being $z^{***}=.031/0.87$. It is easy to see that the range of z^{**} is $(1,\infty)$, and the range of z^{***} is $[0,1)$. It follows that, due to the restriction (8d), two lines are found that cross somewhere in the triangle formed by the points $(x,y)=(0,0),(1,1)$, and $(1,0)$.

We can study boundary solutions for A and B using the unidentified solution in table 2 and the corner points of the area in figure 4. The corner points are (1.000, .356), (1.025, .365), (1.000, -.020) and (1.040, -.021). Using these values for (x,y) we can derive four new sets of parameters, see table 4. In table 4 we get an idea of the ranges of identifiability.

TABLE 4. Solution corresponding with corner points in figure 6.

(x,y)coordinates		1.000	.356	1.025	.365	1.000	-.020	1.040	-.021
Mekranoti	Males	.510	.490	.523	.477	.224	.776	.233	.767
	Females	.534	.466	.548	.452	.262	.738	.272	.728
	Children	.965	.035	.990	.010	.945	.055	.983	.017
Kanela	Males	.641	.359	.657	.343	.432	.568	.449	.551
	Females	.498	.502	.510	.490	.204	.796	.212	.788
	Children	.975	.025	1.000	.000	.961	.039	1.000	.000
Bororo	Males	.706	.294	.723	.277	.534	.466	.555	.445
	Females	.468	.532	.480	.520	.158	.842	.164	.836
	Children	.931	.069	.954	.046	.891	.109	.927	.073
Xavente	Males	.788	.212	.808	.192	.664	.336	.690	.310
	Females	.369	.631	.378	.622	.000	1.000	.000	1.000
	Children	.914	.086	.937	.063	.864	.136	.898	.102
Activities									
	Being idle	.781	.174	.766	.174	.781	.398	.766	.398
	Sleeping	.087	.000	.085	.000	.087	.032	.085	.032
	Caring	.000	.105	.003	.105	.000	.066	.003	.066
	Nonsubsistence	.023	.514	.035	.514	.023	.333	.035	.333
	Domestic	.082	.119	.082	.119	.082	.105	.082	.105
	'Wild'	.028	.089	.029	.089	.028	.066	.029	.066

If there are three latent budgets, things become much more complicated. For $K=2$ we can picture the situation in a two-dimensional space, but for $K=3$ we need a six-dimensional space. Another way to proceed is to use figure 2 and 3 as a starting point. In figure 2 we see how the expected budgets for the rows in Π can be derived as a nonnegative linear combination of the three latent budgets in B using A. We know that the model is not identified, i.e. we can find the same expected budgets for the rows in Π by defining new coordinates in $\tilde{A}$ for the 12 points and new latent budgets in $\tilde{B}$. Graphically this corresponds to making a new triangle. Not every triangle is allowed, firstly because restriction (7c) implies that no row points may fall outside the triangle since the linear combination of the latent budgets should be nonnegative, and secondly, because the corner points cannot be too far away from the twelve row points, because this would result in a violation of restriction (7d). Keeping this in mind we can find an identified solution that has the same fit as the unrestricted solution in the following way.

At least six values have to be fixed in order to identify the model. In order to simplify the interpretation we restrict as much α_{ik}-parameters as possible to zero. Figure 2 suggests that this can be accomplished by setting $\alpha_{1,1}=\alpha_{11,1}=\alpha_{5,2}=\alpha_{6,2}=\alpha_{7,3}=\alpha_{9,3}=0$. So a matrix T can be derived that restricts these values to zero by solving a system of six linear equations (one for each restricted parameter) with six unknowns (the independent values of T). As a second step we have to check whether $B(T^{-1})' \geqslant 0$. For this example this approach works. However, this approach does not always work because $B(T^{-1})'$ can have negative elements. If this happens at least one of the α_{ik}-parameters being restricted to zero has to be restricted to a less extreme value.

To summarize, the model is unidentified, but the identification can be better understood by making plots such as figure 4 (for $K=2$) and in figure 2 and 3 (for $K=3$). If $K>3$, then things become even more complicated. Firstly, for $K=3$ we have to approach the convex hull of the rows points with three lines, and for $K=4$, for example, we have to approach the three dimensional convex hull using four planes. Secondly, it is more difficult to solve the problem that, given some acceptable T for which $AT \geqslant 0$, the matrix $B(T^{-1})'$ can have negative elements. At the moment we are working on these problems.

After having imposed constraints we can check for identification by checking T. If the only admissible T to go from A and B to $\tilde{A}$ and $\tilde{B}$ is the identity matrix, then the model is identified. This method to check for identification was used before, for example, by MOOIJAART (1982).

THEOREM 2. *The number of degrees of freedom of the LBA model is* $(I-K)(J-K)$.

PROOF. The number of degrees of freedom is equal to the number of independent cells minus the number of independent parameters. The number of independent cells is $I(J-1)$. Due to restrictions (7a) and (7b) the number of row parameters is $I(K-1)$, the number of column parameters is $K(J-1)$. However, these parameters are not identified, and therefore not independent. We have to subtract the free elements of the matrix T to find the correct number of independent parameters, i.e. we have to subtract $K(K-1)$. This gives $I(J-1)-[I(K-1)+K(J-1)-K(K-1)]=(I-K)(J-K)$. QED

4. STABILITY

Now that we have an identified solution, we can study the standard errors of the parameters. Thus we have an idea of the stability of the solution. We derive the standard errors using the delta method (see, for example, BISHOP, FIENBERG and HOLLAND, 1975, section 14.6.3).

We first redefine the parameters. Let

$$\alpha_{ik} = \exp \omega_{ik} / \sum_r \exp \omega_{ir} \tag{11}$$

$$\beta_{jk} = \exp \theta_{jk} / \sum_1 \exp \theta_{1k} \tag{12}$$

The partial derivates of α_{ik} and β_{jk} with respect to the new parameters ω_{ik} and θ_{jk} are

$$\partial\alpha_{ik}/\partial\omega_{st} = \delta^{is}\alpha_{st}(\delta^{kt}-\alpha_{sk}) \tag{13}$$

$$\partial\beta_{jk}/\partial\theta_{nm} = \delta^{km}\beta_{nm}(\delta^{jn}-\beta_{jm}) \tag{14}$$

where δ^{is} is a Kronecker delta. Now

$$\partial\pi_{ij}/\partial\omega_{st} = \sum_k \beta_{jk}\partial\alpha_{ik}/\partial\omega_{st} = \delta^{is}\alpha_{st}(\beta_{jt}-\pi_{sj}) \tag{15}$$

$$\partial\pi_{ij}/\partial\theta_{nm} = \sum_k \alpha_{ik}\partial\beta_{jk}/\partial\theta_{nm} = \alpha_{im}\beta_{nm}(\delta^{jn}-\beta_{jm}) \tag{16}$$

We find the asymptotic covariance matrix of the estimators of the parameters ω_{st} and θ_{nm} in the following way. First we collect the partial derivates (15) and (16) in a matrix G having IJ rows (corresponding with π_{ij}) and $IK+JK$ columns, corresponding with the estimators of the parameters ω_{st} and θ_{nm}. The elements of columns of G that correspond with fixed parameters have value zero. Values p_{ij}/n_{i+} are collected in a diagonal matrix D. Then the covariances between the estimators of the parameters ω_{st} and θ_{nm} are the elements of the matrix

$$\Sigma = 1/n_{++}(G'D^{-1}G)^{-1} \tag{17}$$

where from $(G'D^{-1}G)$ the Moore-Penrose generalized inverse is taken.

In order to find the covariances between the parameters α_{ik} and β_{jk}, we collect the partial derivatives (13) and (14) in the matrix W, having α_{ik} and β_{jk} in the rows and ω_{st} and θ_{nm} in the columns. Then the covariance matrix Σ_c for the parameters α_{ik} and β_{jk} is equal to

$$\Sigma_c = W\Sigma W' \tag{18}$$

The square roots of the diagonal elements of Σ_c are the asymptotic standard errors of the parameters α_{ik} and β_{jk}.

As an example we show the identified solution discussed above, where $\alpha_{1,1}=\alpha_{11,1}=\alpha_{5,2}=\alpha_{6,2}=\alpha_{7,3}=\alpha_{9,3}=0$. For $K=3$ latent budgets we find a fit of $G^2=37.0$. The parameter values with their standard errors are reported in table 5.

TABLE 5. Parameter estimates with standard errors for identified solution with $\alpha_{1,1}=\alpha_{11,1}=\alpha_{5,2}=\alpha_{6,2}=\alpha_{7,3}=\alpha_{9,3}=0$.

Row parameters:	Males			Females			Children		
	k=1	k=2	k=3	k=1	k=2	k=3	k=1	k=2	k=3
Mekranoti	.000*	.949	.051	.267	.106	.627	.890	.087	.023
		(.09)	(.09)	(.13)	(.11)	(.15)	(.10)	(.08)	
Kanela	.302	.505	.193	.250	.000*	.750	.955	.000*	.045
	(.13)	(.16)	(.11)	(.15)		(.15)	(.08)		(.08)
Bororo	.289	.711	.000*	.145	.218	.637	.779	.222	.000*
	(.15)	(.15)		(.14)	(.14)	(.17)	(.11)	(.11)	
Xavente	.473	.517	.010	.000*	.226	.774	.801	.116	.074
	(.14)	(.14)	(.05)		(.15)	(.15)	(.10)	(.09)	(.07)

Activities	k=1	k=2	k=3
Being idle	.806	.480	.379
	(.04)	(.05)	(.07)
Sleeping	.093	.039	.032
	(.02)	(.03)	(.02)
Caring	.002	.002	.117
	(.01)	(.01)	(.03)
Nonsubsistence	.014	.242	.357
	(.03)	(.05)	(.06)
Domestic	.084	.095	.111
	(.03)	(.04)	(.05)
'Wild'	.004	.143	.004
	(.01)	(.03)	(.01)

5. CONCLUSION

The latent time-budget model is a tool that can be helpful for obtaining a parsimonious description of time-budgets. Its parameters are interpretable as proportions, which makes communication with non-statisticians much easier. Relations with other methods, as well as the incorporation of row and/or column structure, are studied in DE LEEUW and VAN DER HEIJDEN (1989) and VAN DER HEIJDEN, MOOIJAART and DE LEEUW (1989).

The statistical properties of the model are derived under the assumption that the time-budgets are collected using random spot check methods. If the data are not collected by random spot check methods, but the budgets are formed by aggregating continuous-time event-histories (compare the introduction), then the statistical analysis becomes more complicated. Estimates of the latent budget parameters will still be consistent, but the likelihood function for these marginals is now much more complicated, which implies that the estimates are not longer efficient. In the case of a stationary continuous time Markov chain the asymptotic distribution of these occupancy times has been derived by GOOD (1961), compare also BILLINGSLEY (1961). Unfortunately in this case the asymptotic distribution also involves the transition matrix of the chain.

We will compare LBA model with logcontrast principal component analysis (AITCHISON, 1986). In Aitchison's approach to the statistical analysis of compositional data the first step is to transform the compositional data in Π. One transformation is to replace the values π_{ij} by $z_{ij}=\log(\pi_{ij}/g(\pi_{ij}))$, where $g(\pi_{ij})=(\pi_{i1}...\pi_{iJ})^{1/J}$, i.e. the geometric mean. Elements z_{ij} are collected into the matrix Z. Logcontrast principal component analysis decomposes the covariance matrix Γ derived from Z as $\Gamma=Q\Phi Q'$, with $Q'Q=I$, where Φ is the matrix with eigenvalues in descending order. The component loadings are defined as $Q\Phi^{1/2}$, and the component scores as $P=ZQ$. For our example the component scores are shown in figure 5 and the component loadings in figure 6. The eigenvalues are 2.675, 1.229, .283, .109 and .042, and this shows that the first two dimensions account for 90.0% of the variance.

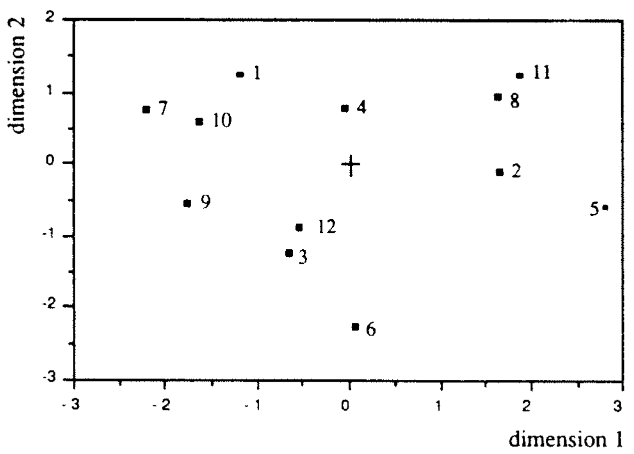

FIGURE 5. Component scores of logcontrast principal component analysis.

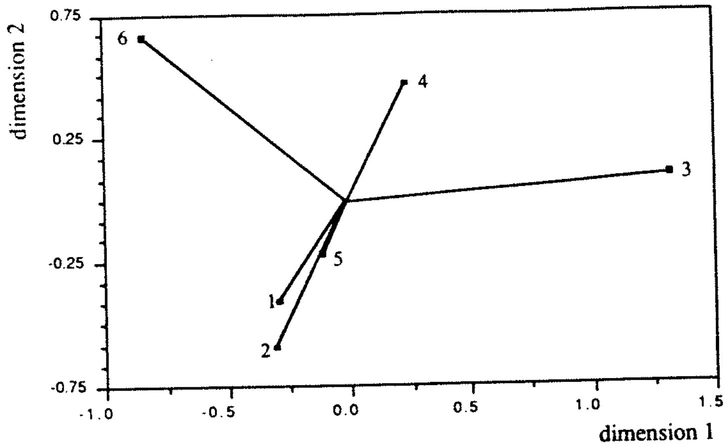

FIGURE 6. Component loadings of logcontrast principal component analysis.

Figures 5 and 6 are strikingly similar to figures 2 and 3. This can be explained as follows. LBA gives a rank 3 approximation of the matrix Π using a nonnegative decomposition. By eliminating the mean vector to the simplex two-dimensional representations are obtained in figures 2 and 3. In logcontrast principal component analysis the matrix Π is transformed into a matrix Z; this transformation is monotonous for each row. This transformation has the result that elements z_{ij} of the matrix Z have the property that $z_{i+}=0$. As a second step Z is transformed into $\tilde{Z}$, that is in deviation from column means, and therefore $\tilde{z}_{i+}=0$ and $\tilde{z}_{+j}=0$. So the rank $\tilde{Z}$ is reduced with 1 to min $(I-1, J-1)$. Then $\tilde{Z}$ is decomposed as $\tilde{Z}=P\Phi^{1/2}Q'$, the first two dimensions are displayed in figure 5 and 6. So it is not surprising that LBA and logcontrast principal component analysis give similar graphical representations: LBA gives a nonnegative reduced rank approximation of Π, and logcontrast principal component analysis gives a reduced rank approximation of a transformation of Π.

LBA has the advantage that its parameters are very easily interpretable. We think that LBA is preferable over logcontrast principal component analysis if the data are collected with the random spot check method, because then it is allowed to draw conclusions about the populations using sample data. For logcontrast principal component analysis such conclusions are allowed if we assume that the rows of Z are replications drawn from a multinormal distribution. Then the covariance matrix Γ can be decomposed using confirmatory factor analysis, for example. For the data in table 1 we think that it is not very attractive to consider the rows of the matrix (the groups) as replications.

ACKNOWLEDGEMENT

We are grateful for helpful comments of the editor and two referees. This research project was performed while the second author worked for the Department of Psychometrics, University of Leiden.

REFERENCES

AITCHISON, J. (1986), *The statistical analysis of compositional data*, Chapman and Hall, London.

BILLINGSLEY, P. (1961), Statistical methods in Markov chains, *Annals of Mathematical Statistics 32,* 12-40.

BISHOP, Y.M.M., S.E. FIENBERG and P.W. HOLLAND (1975), *Discrete multivariate analysis,* M.I.T.-Press, Cambridge.

DE LEEUW, J., P.G.M. VAN DER HEIJDEN and I. KREFT (1985), Homogeneity analysis of event history data, *Methods of Operations Research 50,* 299-316.

DE LEEUW, J. and P.G.M. VAN DER HEIJDEN (1988), The analysis of time-budgets with a latent-time-budget model, in E. Diday et al. (eds.), *Data analysis and informatics 5,* North-Holland, Amsterdam, 159-166.

DE LEEUW, J. and P.G.M. VAN DER HEIJDEN (1989), Reduced rank models for contingency tables, Department of Psychology, Leiden (PRM 06-89).

DEMPSTER, A.P., N.M. LAIRD and D.B. RUBIN 1977, Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society, Series B 39,* 1-38.

ELLIOTT, D.H. (1984), Statistical analysis of time-budget data, in: A.S. Harvey, A. Szalai, D.H. Elliott, P.J. Stone and S.M. Clark, *Time-budget research,* Campus Verlag, New York, 84-115.

GOOD, I.J. (1961), The frequency count of a Markov chain and the transition to continuous time, *Annals of Mathematical Statistics 32,* 41-48.

GOODMAN, L.A. (1985), The analysis of cross-classified data having ordered and/or unordered categories: association models, correlation models, and asymmetry models for contingency tables with or without missing entries, *Annals of Statistics 13,* 10-69.

GOODMAN, L.A. (1986), Some useful extensions of the usual correspondence analysis approach and the usual log-linear models approach in the analysis of contingency tables, *International Statistical Review 54,* 243-309.

GREENACRE, M.J. (1984), *Theory and applications of correspondence analysis,* Academic Press, New York.

GROSS, D.R. (1984), Time allocation: a tool for the study of cultural behavior, *Annual Review of Anthropology 13,* 519-558.

GROSS, D.R., J. RUBIN and N.M. FLOWERS (1985), All in a day's time: random spot check data can describe the texture of a day's activities, Presented at the Annual Meeting of the American Association for the Advancement of Science, Los Angeles, May 29, 1985.

HARVEY, A.S., A. SZALAI, D.H. ELLIOTT, P.J. STONE and S.M. CLARK (1984), *Time-budget research,* Campus Verlag, New York.

MOOIJAART, A. (1982), Latent structure analysis for categorical variables, in: K.G. JÖRESKOG and H. WOLD (1982), *Systems under indirect observation,* North-Holland, Amsterdam, 1-18

PARKES, D.N. and N.J. THRIFT (1980), *Times, spaces and places,* Wiley, New York.

ROBINSON, J.P. (1985), The validity and reliability of diaries versus alternative time use measures, in: F.T. JUSTER and F.P. STAFFORD (eds), *Time, goods & well-being,* University of Michigan, Michigan, 33-62.

SAPORTA, G. (1981), Méthodes exploratoires d'analyse de données temporelles, *Cahiers du Buro no 37-38,* Paris.

SAPORTA, G. (1985), Data analysis for numerical and categorical individual time-series, *Applied Stochastic Models and Data Analysis 1,* 109-119.

STAIKOV, Z. (ed.) (1982), *It's about time,* Bulgarian Academy of Sciences, Sofia.

SZALAI, A. (ed.) (1972), *The use of time,* Mouton, The Hague.

VAN DER HEIJDEN, P.G.M. (1987), *Correspondence analysis of longitudinal categorical data,* D.S.W.O.-Press, Leiden.

VAN DER HEIJDEN, P.G.M. and J. DE LEEUW (1989), Correspondence analysis, with special attention to the analysis of panel data and event history data, in: C.C. CLOGG (ed.), (1989), *Sociological Methodology 1989,* Basil Blackwell, Oxford, pp. 43-87.

VAN DER HEIJDEN, P.G.M., A. MOOIJAART and J. DE LEEUW (1989), Latent budget analysis, in: A. DECARLI, B.J. FRANCIS, R. GILCHRIST and G.U.H. SEEBER (eds.) (1989), *Statistical modelling, Proceedings, Trento 1989,* Springer Verlag, Berlin, 301-313.

Received March 1989, Revised September 1989.