The analysis of time-budgets with a latent time-budget model

Jan de Leeuw* & Peter G.M. van der Heijden**

* Departments of Data Theory,
Faculty of Social Sciences,
University of Leiden,
Middelstegracht 4,
2312 TW Leiden,
The Netherlands

** Department of Methodology,
Faculty of Psychology,
University of Leiden,
Hooigracht 15,
2312 KM Leiden,
The Netherlands

1. Introduction

Event history data are data specifying the sequence and times in which objects or individuals are in specific states (categories) during some period of time, where the number of states is finite. These data are collected in for example sociology, anthropology, ethology and psychology in studies concerning the use of time by humans and non-humans. For each specific time point, the state each object is in can be derived from event history data. Theoretically we are working in continuous time, and the number of time-points during some time-period goes to infinity. In practice, however, time is always measured in discrete units (for example months, days, minutes), and thus the situation simplifies for data analysis purposes. Aggregated forms of event history data are also known as **time-budget data**, or **time-allocation data**.

In sociology data are often collected in the form of **diaries**. A very valuable review of contemporary model-based data analysis techniques for the analysis of diaries and other 'social dynamics' data is Tuma and Hannan (1984). These techniques require a great deal of prior knowledge, and can be applied only in a fairly restrictive class of situations (very many observations, and only very few states).

In ethology attention for time allocation seems to be growing. We mention here Bernstein(1972), Barash (1974), Barnard et al. (1984), Boy & Duncan (1979), Duncan (1980, 1985) and Arnold and Trillmich (1985) as examples. It will be clear that, unlike in sociology, it is impossible to work with diaries in ethology; the behavior is observed by the researcher. From a methodological point of view applications in ethology are somewhat special, because ethologists have a lot of freedom to choose their population samples and time samples. There are virtually no limits on the availability and the cooperativeness of sticklebacks and wolf spiders. A comparison of the different time sampling methods used by ethologists is given by Altman (1974) and Slater (1973).

In anthropology time allocation studies also seem to become more popular. In a recent paper Gross (1984) reviews the use of time allocation methods in anthropology. In studies using time allocation methods the investigators collect information on how the individuals in a community or group spend their time. Gross reviews many applications of time allocation studies, and comments extensively on their practical usefulness and their methodological aspects. His conclusion is worth repeating. "Time allocation techniques comprise a valuable tool for recording and analyzing human behavior in a natural context. The power and usefulness of time allocation techniques have been demonstrated in village level studies and national and international samples. There are still many problems to be solved to achieve maximum efficiency, accuracy, and comparability of data sets." (Gross, 1984, p. 548). Gross compares these various methods in terms of their usefulness for anthropologists, who obviously have to deal with practical constraints that are quite different from those of ethologists. It appears from Gross' evaluation that the **random spot check method** is the preferable one. The studies of Erasmus (1955) and Johnson (1975) are given as major examples of this approach.

In spot check methods investigators take 'snapshot-like' recordings of behavior. The idea is that, if a person is baking bread 7 times out of 100 times that we have made a spot-check of his home, then we assume that he spends approximately 7% of his time baking bread. Clearly there are various practical and theoretical limitations to this method, which almost immediately come to mind. They are discussed in detail by Gross (1985, p. 537-546), but they do not alter the final conclusion about the usefulness of the method. There is an appealing modern illustration of the method in Gross et al. (1985). Behavior was sampled, between 6:00 A.M. and 8:00 P.M., in four different South American Indian Villages, located in Central Brazil. We shall use this example to illustrate possible analytic techniques for spot-check (and other time allocation) data.

2. Ways to analyze time-budget data

If we look at the ways time allocation data are usually analyzed, we find mostly purely descriptive and tabulatory techniques. In his review paper Gross (1984, p. 546-548) also has no suggestions about how one should proceed beyond the purely descriptive point. There are much more specific proposals for the analysis of time budgets in the publications of the group involved in the Multinational Time Budget Project (Szalai, 1972). Converse (1972) pioneered the use of multivariate analysis techniques on such data, and this was subsequently taken up by Clark et al. (1982), Harvey et al. (1984). The multivariate analysis techniques that are used, however, are standard packaged techniques that do not take special properties of time allocation data into account. This makes their use tentative, at best.

As an example in the context of the Multinational Time Budget Project, Stone (1972) used Smallest Space Analysis, a multidimensional scaling technique, in order to obtain a multidimensional representation of the 15 cities taking part in the project. In order to be able to do this, he computed similarity measures between the cities on the basis of their time budgets. Other similarity measures are to be found in Harvey (1984). These measures emphasize the differences between the better filled categories. This will often be a drawback, especially when we are also interested in differences in the smaller categories.

Another proposal for the analysis of time-budget data is to use correspondence analysis (Gifi, 1981; Greenacre, 1984). Correspondence analysis can be described as a technique with which a geometrical representation of rows and columns can be obtained. Distances between either rows or columns approximate chi-square distances in Euclidean space. Chi-square distances correct for different margins of the categories, and therefore solutions are not necessarily dominated by the better filled categories. The number of applications of correspondence analysis is growing in the context of time-budget data. Some references are Deville and Saporta (1983), Duncan (1985), De Leeuw et al. (1985), and Van der Heijden (1987). Jambu and Lebeaux (1983) use a data set from the Multinational Time Budget Project as one of the examples throughout their book.

Although correspondence analysis seems to provide us with an appropriate measure for differences between time budgets, one important problem is not solved. Clark et al. (1982) formulate this as follows: "Multivariate techniques are very ill-suited to serve as communications devices between the community of scientific researchers and the larger society. Indeed, the techniques are not even particular appropriate for communicating ideas within many parts of the scientific community." (l.c., p. 69). In this paper we shall discuss a special purpose multivariate technique for the analysis of spot-check and other time-budget data. It is our opinion that the type of representation derived by this technique can be communicated very well, both inside and outside the scientific community. The representations are much more economical than long lists of tables or descriptive diagrams, and they emphasize the most important variation in the data. Thus we circumvent at least some of the disadvantages of tabular analysis (Hirshi and Selvin, 1973; Gifi, 1981). We hope to illustrate the successfulness of our technique by analyzing the example presented by Gross et al. (1985). We will compare this technique with correspondence analysis, both theoretically as well as empirically.

3. A latent time-budget model

Suppose an individual or group is engaged, during a period T, in any one of m different activities. The basic data in this paper are measurements of the distribution of time over the different activities. It is the business of the scientist to define, as carefully as possible, the population of individuals he is interested in, the period over which he intends to study his population, and the classification of activities he intends to use. Compare Gross (1984, p. 537-546), for instance, for considerations which must be taken into account in anthropology. Of course the classification of the possible activities also plays a major role in the design of time budget studies in other fields.

The type of data that we consider in this paper does not take the time of day into account. In anthopological terms it cannot give us the 'texture of the day' of the various groups or individuals in the study, it can only provide us with an idea how the various activities are distributed over individuals or groups. If the groups are men and women in a particular culture, for instance, we can see which activities belong to the task of the typical woman, and which to the task of the typical man,but we cannot see how the available time is used to plan and execute activities. This format is used in the older time-budget studies, comparing unmarried women in Yugoslavia with married men in the U.S.A., and so on. Groups are defined by crossing properties such as sex, country, and marital status. In the anthropological example that we shall use, the groups are obtained by crossing the societies of the Mekranoti, Xavante, Kanela, and Bororo Indians with the age-sex variable adult males, adult females, and juveniles under 15. Thus their are twelve groups.

Another piece of information that we loose if we go from event history to time budget data is the sequence of activities. We aggregate over time, and the sequence, or the count of the transitions, simply gets lost. Thus event history data are inherently richer, but this is also their weakness. As always rich data structures can easily lead to many empty cells, and to overparametrization. For this reason the time budgets, which are as it were a marginal of the event history data matrix, are more robust data.

The data can be collected in an n x m matrix, where n is the number of groups or individuals and m the number of activities. The entries of the matrix are integers n_{ij}, which is the number of individuals in group i that were engaged in activity j during random spot-checks, or the number of times individual i was observed doing j. The total number of spot-checks for row i is n_{i+}, which is supposed to be a fixed number, determined by the design of the experiment. We have decided, beforehand, that we are going to check in on the Mekranoti families, say, 75 times. Because the family sizes are fixed during the experiment, this means, for example, that we have 18 x 75 observations on adult males, and 94 x 75 observations on juveniles. Of course it is possible that various things go wrong during the experiment (people may be out hunting, persons may die, and so on), but this does not change the fact that essentially the n_{i+} are fixed. But the n_{ij} are a different matter. They are the outcomes of the experiment, and it is best to conceptualize them as random variables. If we repeat the experiment, with the same n_{i+}, we shall undoubtedly find somewhat different n_{ij}. It is clear that $p_{ij} = n_{ij}/n_{i+}$ can be used as an estimate of π_{ij}, the proportion of time spent by a typical member of group i on activity j, or the proportion of time spent by the individual i at activity j on a typical day.

If we assume independence between observations the n_{ij} in row i are multinomially distributed with means $E(n_{ij}) = n_{i+}\pi_{ij}$. A first result that interests us, as a sort of baseline, is whether the π_{ij} are different for the different groups. Of course we expect that they will be very different indeed, otherwise our categories of behavior or our groups of individuals must have been defined in a rather uninteresting way. The usual test for equi-distribution in a rectangular table is the chi-square test.

Here we discuss a model for the analysis of time budgets, which can be considered to be a specially adapted form of factor analysis. We assume that the equidistribution model is untenable, and we analyze the difference in the distributions for the various i. Remember that the theoretical time budgets are given by π_{ij}, where i = 1, ... , n, j = 1, ... , m, and where $\Sigma_{j=1}^{m} \pi_{ij} = 1$ for all i. The model is

$$\pi_{ij} = \Sigma_{k=1}^{p} \beta_{ik}\alpha_{jk}, \tag{1}$$

with restrictions $\beta_{ik} \geq 0$, $\alpha_{jk} \geq 0$, and $\Sigma_{k=1}^{p} \beta_{ik} = 1 = \Sigma_{j=1}^{m} \alpha_{jk}$. The number of degrees of freedom can be computed as the number of independent cells minus the number of independent parameters, being n(m - 1) - {p(m - 1) + (p - 1)n}. This model can be interpreted as a model describing how a theoretical time budget in row i is the result of p latent time budgets, given by α_{jk}, on which it 'loads'. Each theoretical as well as each latent time budget sums to 1. Values β_{ik} show for which proportions the theoretical time budget of row i is made up from latent time budget k. The number of latent time budgets has to be specified by the researcher. In case of p = 1, model (1) is equal to the usual independence model, having n(m - 1) - (m - 1) = (n - 1)(m - 1) degrees of freedom.

In case we postulate that our time budgets are sampled under a product multinomial distribution, maximum likelihood estimation can be accomplished using the EM algorithm (Dempster, Laird, and Rubin, 1977).

Maximum likelihood estimation

The logarithm of the likelihood is, except for an irrelevant constant,

$$\mathcal{L}(\alpha,\beta) = \Sigma_{i=1}^{n}\Sigma_{j=1}^{m} n_{ij} \ln \Sigma_{k=1}^{p} \beta_{ik}\alpha_{jk}. \tag{2}$$

This must be maximized over the unknowns, with restrictions $\beta_{ik} \geq 0$, $\alpha_{jk} \geq 0$, and $\Sigma_{k=1}^{p} \beta_{ik} = 1 = \Sigma_{j=1}^{m} \alpha_{jk}$. We present an elementary derivation of an EM-type algorithm for this problem.

Suppose $\underline{\beta}_{ik}$ and $\underline{\alpha}_{jk}$ are the current best estimates at some point during the iterations of the algorithm, giving theoretical values $\underline{\pi}_{ij}$ according to (1). Also define $n_{ijk} = n_{ij}\underline{\alpha}_{jk}\underline{\beta}_{ik}/\underline{\pi}_{ij}$. We use the notational convention of replacing an index over which we have summed by a plus. Observe that $n_{ij+} = n_{ij}$, and the n_{ijk} can be seen as distributing the observed budgets over the p dimensions.

Theorem 1. Consider the algorithm which computes updates by the rules $\alpha_{jk}^{+} = n_{+jk}/n_{++k}$ and $\beta_{ik}^{+} = n_{i+k}/n_{i++}$. Then $\mathcal{L}(\alpha^{+},\beta^{+}) \geq \mathcal{L}(\underline{\alpha},\underline{\beta})$.

Proof. From the concavity of the logarithm,

$$\ln \pi_{ij}/\underline{\pi}_{ij} = \ln \{\Sigma_{k=1}^{p} \underline{\alpha}_{jk}\underline{\beta}_{ik}(\alpha_{jk}\beta_{ik}/\underline{\alpha}_{jk}\underline{\beta}_{ik})/\underline{\pi}_{ij}\} \geq$$

$$\geq \{\Sigma_{k=1}^{p} \underline{\alpha}_{jk}\underline{\beta}_{ik} \ln(\alpha_{jk}\beta_{ik}/\underline{\alpha}_{jk}\underline{\beta}_{ik})\}/\underline{\pi}_{ij}. \tag{3}$$

Substitution in (2), using (3), gives

$$\mathcal{L}(\alpha,\beta) \geq \mathcal{L}(\underline{\alpha},\underline{\beta}) + \mathcal{D}(\alpha,\beta,\underline{\alpha},\underline{\beta}), \tag{4}$$

where

$$\mathcal{D}(\alpha,\beta,\underline{\alpha},\underline{\beta}) = \Sigma_{i=1}^{n}\Sigma_{j=1}^{m}\Sigma_{k=1}^{p} n_{ijk} \ln \alpha_{jk}\beta_{ik} - \Sigma_{i=1}^{n}\Sigma_{j=1}^{m}\Sigma_{k=1}^{p} n_{ijk} \ln \underline{\alpha}_{jk}\underline{\beta}_{ik}. \tag{5}$$

Moreover (4) is an equality if $(\alpha,\beta) = (\underline{\alpha},\underline{\beta})$. It is easy to see that we find (α^+,β^+), the successor of $(\underline{\alpha},\underline{\beta})$, by maximizing $\mathcal{D}(\alpha,\beta,\underline{\alpha},\underline{\beta})$ over (α,β). Thus

$$\mathcal{L}(\alpha^+,\beta^+) \geq \mathcal{L}(\underline{\alpha},\underline{\beta}) + \mathcal{D}(\alpha^+,\beta^+,\underline{\alpha},\underline{\beta}) \geq$$

$$\geq \mathcal{L}(\underline{\alpha},\underline{\beta}) + \mathcal{D}(\underline{\alpha},\underline{\beta},\underline{\alpha},\underline{\beta}) = \mathcal{L}(\underline{\alpha},\underline{\beta}). \tag{6}$$

QED.

Thus the algorithm increases the likelihood in each step. In fact if we build in the rule that the algorithm stops if $(\underline{\alpha},\underline{\beta})$ already maximizes $\mathcal{D}(\alpha,\beta,\underline{\alpha},\underline{\beta})$, then either the algorithm stops at a stationary point where the likelihood equations are satisfied, or it generates an infinite sequence for which the increase in the likelihood is strict. The sequence of solutions that is computed has accumulation points, because the restrictions define a compact set, and each accumulation point has the same value of the likelihood and satisfies the likelihood equations. This means that we can say for all practical purposes that the algorithm converges.

The stationary points of the algorithm have an interesting property, which we prove next. It turns out that the matrices of observed and expected values have the same marginals. This is obvious for the row marginals, but for the columns the situation is a bit more complex.

Theorem 2. At a stationary point $\Sigma_{i=1}^{n} n_{i+}\pi_{ij} = n_{+j}$.

Proof. At a stationary point of the algorithm we have

$$\Sigma_{j=1}^{m} (n_{ij}/\pi_{ij})\alpha_{jk} = n_{i+}, \tag{7a}$$

$$\Sigma_{i=1}^{n} (n_{ij}/\pi_{ij})\beta_{ik} = \mu_k, \tag{7b}$$

with μ_k Lagrange multipliers. If we multiply both sides of (7b) by α_{jk}, and sum over k, we find

$$n_{+j} = \Sigma_{k=1}^{p} \mu_k\alpha_{jk}. \tag{8}$$

If we multiply both sides of (7b) by α_{jk}, sum over j, and use (7a), we find

$$\Sigma_{i=1}^{n} n_{i+}\beta_{ik} = \mu_k. \tag{9}$$

Now multiply both sides of (9) by α_{jk}, and sum over k. This gives, using (8),

$$\Sigma_{i=1}^{n} n_{i+}\pi_{ij} = \Sigma_{k=1}^{p} \mu_k\alpha_{jk} = n_{+j}. \tag{10}$$

QED.

After the maximization of the likelihood is carried out, the difference between the observed and expected time budgets can be tested using chi-square statistics. However, this test could be problematic, since the asymptotic distribution only holds if in a specific time budget each observation has the same theoretical distribution, and subsequent observations are independent. The latter assumption will often be violated because activities of different objects will not be independent: for example, a mother is cooking, the child helps her. This dependence can be taken care of, for instance by noting only the behavior of one of the objects, for example, the mother. We find another type of violation in case the n_{ij} are derived from event histories, and a frequency corresponds to for example a minute spent on some activity. Clearly activities often take longer than a minute, and in this case the dependence of subsequent observations is considerable. So in this situation model (1) should only be used as a descriptive tool. However, violation of this assumption does not have to be severe in case of data sampling using the random spot check method discussed in the introduction. Here, by making the intervals between subsequent spot checks large, the dependence between the observations can probably be made small. So in this situation, model (1) can be used for inferential purposes.

The objective of the latent budget model will be clear: it aims at a sparse description of the data in terms of typical time budgets, provided by α_{jk}. Values β_{ik} show how (groups of) objects load on these typical time budgets. Model (1) seems to be very well suited for the analysis of time budgets due to its restrictions that loadings β_{ik} and estimated proportions α_{jk} should be larger than zero and add up to one over time budgets and categories, respectively.

It is quite easy to give a generalization of model (1) to aggregated event histories. Consider for instance theoretical event histories π_{ijt}, where t denotes time periods. We assume $\Sigma_{j=1}^{m} \pi_{ijt} = 1$, so for each time period t we have theoretical time budgets indexed by i which sum to 1. Now a possible generalization of (1) is

$$\pi_{ijt} = \Sigma_{k=1}^{p} \beta_{ik}\alpha_{jtk}, \tag{11}$$

with restrictions $\beta_{ik} \geq 0$, $\alpha_{jtk} \geq 0$, and $\Sigma_{k=1}^{p} \beta_{ik} = 1 = \Sigma_{j=1}^{m} \alpha_{jtk}$. Model (11) specifies that group or object i has k aggregated latent event histories, where the k'th latent event history is built up of different latent time budgets for each t. Such a model will often be reasonable, because in most applications it will be probable that in different time periods the latent time budgets differ. So for some k, values α_{jtk} will provide us with an aggregated latent event history. Of course, the remarks made above for the points of inference and description also hold here: actually model (11) can only be used for inferential purposes in case the data are sampled using the random spot check method, where observations are made as much as possible at distinct times, so that the dependence of observations becomes negligible.

4. Relations with correspondence analysis

The resemblance between the latent budget model and correspondence analysis is large, in the sense that the approximation of π_{ij} provided by a t-dimensional correspondence analysis solution will be often about the same as the approximation provided by the model in case of (t + 1) latent time budgets. Intuitively this can be made clear by realizing that in a two-dimensional correspondence analysis plot the location of each profile point (observed time budget) can be expressed as a weighted average of three typical time budgets placed at the periphery of the cloud of profile points. Another way to make this clear is by comparing model (1) with the model fitted by correspondence analysis, which is

$$\pi_{ij} = \pi_{i.}\pi_{.j}(1 + \Sigma_{s=1}^{t} r_{is}c_{js}\lambda_s), \tag{12}$$

where s is an index for the dimension. Of course correspondence analysis is usually presented as a geometric technique, without any model being involved. But the interpretation of the technique as a method to fit model (12) is valuable as well (Goodman, 1985).

In case $t = 0$, and $p = 1$, the approximation of both model (1) and (12) is equal to the independence model approximation. In case of $t = 1$ and $p = 2$, (1) as well as (12) approximate π_{ij} using the sum of two products of a row term and a column term. In general, in case of $t = q$ and $p = (q + 1)$, both in (1) and (12) π_{ij} is approximated using the sum of $q + 1$ products of a row and a column term. Although the approximations given by (1) and (12) are often about the same, this is not necessarily the case. First of all this is due to the fact that the restrictions on parameters β_{ik} and α_{jk} are different from those on the scores r_{is} and c_{js}. Secondly, in case of model (1), maximizing the likelihood is asymptotically equivalent to minimizing the value of the chi-square statistic

$$S = n_{++} \Sigma_{i=1}^{n}\Sigma_{j=1}^{m}(p_{ij} - \pi_{ij})^2/ \pi_{ij} \tag{13a}$$

Thus the estimates computed by the EM-algorithm are efficient if the model is true. This is not the case for correspondence analysis, where

$$S = n_{++} \Sigma_{i=1}^{n}\Sigma_{j=1}^{m} (p_{ij} - \pi_{ij})^2/(p_{i+}p_{+j}) \tag{13b}$$

is minimized.

The fact that approximations (1) and (12) are nearly the same for this example (although scores r_{is} and c_{js} are quite different from β_{ik} and a_{jk} for all k and s) can further be illustrated when we use the generalization of correspondence analysis proposed by Escofier (1983), see also Van der Heijden & De Leeuw (1985), Van der Heijden (1987). This generalization of correspondence analysis can be written in model form as

$$\pi_{ij} = \vec{\pi}_{ij} + \pi_{i.}\pi_{.j}\Sigma_{s=1}^{t}(r_{is}c_{js}\lambda_s), \tag{14}$$

where $\vec{\pi}_{ij}$ is the expected frequency for cell (i,j) under some model. The generalization is particulary attractive from a data analysis point of view in case $p_{i+} = \vec{\pi}_{i+}$ and $p_{+j} = \vec{\pi}_{+j}$. If the model is (1), then in Theorem 2 it was shown that, due to ML-estimation, indeed $p_{i+} = \vec{\pi}_{i+}$ and $p_{+j} = \vec{\pi}_{+j}$.

5 An example

Gross et al. (1985) recently presented an analysis of random spot check data (compare the introduction) of males, females and kids in four tribes of Amazone Indians. They showed 12 figures (one for each of the three types of persons in each of the four tribes). In these figures for 7 two-hour periods 7 vertical bars display the proportion of time spent in 6 behavior categories, by subdividing each bar into six parts (one part for each category), the length of which represents the proportion of time spent into some category. By measuring the proportions in these bars we could derive a data block with elements n_{ijt} representing the proportion of time that group i spends in category j in time period t. The tribes are the Mekranoti, the Kanela, the Bororo and the Xavante. The six behavior states are 'idle', 'sleep', 'care', 'nonsubst', 'domestic' and 'wild'. For a description of these tribes and a definition of behaviors we refer to Gross et al. (1979, 1985) and Werner et al. (1979). The seven two-hour periods start at 6 A.M. and end at 8 P.M. In our analyses we will add up over the time-periods, in order to facilitate the interpretation. However, as discussed above, this is by no means necessary from a methodological point of view. We will analyze the data matrix, in which the tribes are coded interactively with the types (males, females, kids).

In Table 1 we find estimates for the independence model (row and column margins) in the first column. Table 1 also shows we find the estimates of β_{ik} and α_{jk} for the model with two latent budgets (columns 2 and 3), and for the model with three latent budgets (columns 4, 5, and 6). Considering the results for $p = 2$, we see that the first latent budget is the budget for the adults (for the females predominantly), being less idle and asleep than the marginal time budget, but performing more 'care', 'nonsubst' and 'wild' behavior; the second latent budget is that for the kids, being more idle and asleep, and doing the other behaviors less. Columns 4, 5, and 6 show us results for model (1) with $p = 3$ latent budgets. These latent budgets correspond roughly with the three types of persons, namely the males, the kids and the females respectively.

Because the raw data were not available, we could only compute chi-square measures over the proportions. For the proportions these measures are 2.52 for $p = 1$ (df is 55), .96 for $p = 2$ (df is 38) and .37 for $p = 3$ (df is 21). Consider the case that for each row $n_{i+} = 50$, then the chi-squares are 124, 46 and18 respectively. Comparing these chi-squares with the number of degrees of freedom shows that we cannot conclude for these data that more than three latent time-budgets are necessary to 'explain' the data.

The singular values from correspondence analysis are .355, .224, .155, .068, and .039, which explain 61%, 24%, 12%, 2%, and 1% of the total inertia. The first two dimensions of the solution are shown in Figure 1. Not surprisingly, we see again the three clusters of males, females, and kids.

We can now also use (14) to decompose residuals from model (1) for different numbers of latent time-budgets p. In case of $p = 2$, we find singular values .230, .155, .070, .047 and .016; in case of $p = 3$ these values become .155, .069, .044, .019, .016. This illustrates that for this example approximation (1) in case of p, is about equal to approximation (12) in case of $t = p - 1$.

6 Conclusion and discussion

We have illustrated the latent time budget model, and compared it with correspondence analysis results. We think we can conclude that the model might be useful as a descriptive tool in case it makes sense to think that the observed time budgets are generated by some typical latent ones. This was the case in the example shown. In principle the model can also be used for inferential purposes, whereas correspondence analysis does not (directly) provide this possibility. On the other hand, correspondence analysis provides us with plots that allow a fast interpretation.

Another aspect in comparing correspondence analysis and these models is the following. The latter are perhaps easier to explain to lay men, as a method for finding "typical" time-budgets. On the other hand, it will sometimes be somewhat artificial to think of the observed time budgets as stemming from typical ones, although it was certainly useful here, because the first two dimensions of correspondence analysis showed three clusters of time-budget points. The EM-algorithm for the time budget model can, in many cases, be considered as a form of correspondence analysis in which the dimensions are rotated in such a way that an interpretation in terms of latent budgets becomes possible.

References

Altman, J. (1974). Observational study of behaviour: sampling methods. **Behaviour**, 49, 227-267.

Arnold, W. & Trillmich, F. (1985). Time budgets in galapagos fur seal pups: the influence of the mother's presence and absence on pup activity and play. **Behaviour**, 92, 302-321.

Barash, D.P. (1974). Mother-infant relations in captive woodchucks. **Animal Behaviour**, 22, 446-448.

Barnard, C.J., Thompson, D.B.A., & Stephens, H. (1984). Time budgets, feeding efficiency and flock dynamics in mixed species flocks of lapwings, golden plovers and gulls. **Behaviour**, 80, 44-70.

Bernstein, I.S. (1971). Activity profiles of primate groups. In Scrier, A.M., and Stollnitz, F. (eds.), **Behavior of nonhuman primates.** New york, Academic Press.

Boy, V. & Duncan, P. (1979). Time-budgets of camargue horses. Developmental changes in the time-budgets of foals. **Behaviour**, 71, 187-202.

Clark, S.M., Elliot, D.H., & Harvey, A.S. (1982) Hypercodes and composite variables. In Z. Staikov (ed.), **It's about time.** Sofia, Bulgarian Academy of Sciences.

Converse, P. E. (1972). Country differences in time use. In A. Szalai (ed.), **The use of time.** The Hague, Mouton.

Dempster, A.P., Laird, N.M. & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. **Journal of the Royal Statistical Society**, Series B 39, 1-38.

Deville, J.-C. & Saporta, G. (1983). Correspondence analysis, with an extension towards nominal time series. **Journal of Econometrics**, 22, 169-189.

Duncan, P. (1980). Time-budgets of camargue horses. II. Time-budgets of adult horses and weaned sub-adults. **Behaviour**, 72, 26-49.

Duncan, P. (1985). Time-budgets of camargue horses. III. Environmental influences. **Behaviour**, 77, 188-208.

Erasmus, C.J. (1955). Work patterns in a Mayo Village. **American Anthropology**, 57, 322-333.

Escofier, B. (1983). Analyse de la difference entre deux mesures sur le produit de deux mes ensembles. **Cahiers de l'analyse des données**, 8, 325-329.

Gifi, A. (1981). **Non-linear multivariate analysis**. Leiden: Department of Datatheorie.

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.

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

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

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

Harvey, A.S., Szalai, A., Elliot, D.H., Stone, P.J., & Clarke, S.M. (1984). **Time budget research**. Frankfurt, Campus Verlag.

Heijden, P.G.M. van der (1987). **Correspondence analysis of longitudinal categorical data**. Leiden: DSWO-press.

Heijden, P.G.M. van der & Leeuw, J. de (1985). Correspondence analysis used complementary to loglinear analysis. **Psychometrika**, 50, 429-447.

Hirschi, T. & Selvin, H.C. (1973). **Principles of survey analysis**. Glencoe: Free Press.
Jambu , M. & Lebeaux, M.O. (1983) **Cluster analysis and data analysis.** Amsterdam, North Holland.
Johnson, A. (1975). Time allocation in a Machiguenga community. **Ethnology**, 14, 310-321.
Leeuw, J. de, Heijden, P.G.M. van der, & Kreft, I. (1985). Homogeneity analysis of event history data. **Methods of Operations Research**, 50, 299-316.
Slater, P.J.B. (1976). Describing sequences of behavior. In: P.P.G.Bateson & P.H. Klopfer (eds.). **Perspectives in ethology.** New York.
Stone, P.J. (1972). The analysis of time budget data. In A. Szalai (ed.), **The use of time**. The Hague: Mouton.
Szalai, A. et al. (1972). **The use of time**. The Hague: Mouton.
Tuma, N.B. & Hannan, M.T. (1984). **Social dynamics. Models and methods**. New York: Academic Press.
Werner, D. et al. (1979). Subsistence productivity and hunting effort in native South America. **Human Ecology**, 7, 303-315.

Table 1: Parameter estimates for the time-budget model, for different values of p.

Table 1a: Budget weights, rowwise.

		p=1	p=2		p=3		
		k=1	k=1	k=2	k=1	k=2	k=3
Mekranoti	M	1.000	.763	.237	.107	.832	.061
	F	1.000	.725	.275	.253	.109	.638
	K	1.000	.054	.946	.876	.084	.040
Kanela	M	1.000	.558	.442	.346	.448	.206
	F	1.000	.782	.218	.220	.019	.761
	K	1.000	.038	.962	.929	.009	.063
Bororo	M	1.000	.458	.542	.363	.625	.012
	F	1.000	.828	.172	.146	.207	.647
	K	1.000	.108	.892	.783	.200	.017
Xavente	M	1.000	.331	.669	.520	.457	.024
	F	1.000	.982	.018	.002	.216	.783
	K	1.000	.134	.866	.799	.110	.091

Table 1b: latent budgets, columnwise.

Idle	.594	.391	.781	.817	.437	.391
Sleep	.060	.031	.087	.095	.032	.034
Care	.032	.068	.000	.000	.000	.116
Nonsubst	.174	.338	.023	.005	.271	.348
Domestic	.093	.105	.081	.080	.096	.110
Wild	.047	.067	.028	.003	.163	.000
Fit:	2.519	.963		.370		
Df:	55	38		21		
Chi-square ($n_{i+} = 50$)	124	46		18		

Figure 1: Activities (6) are large, Tribes x Type of person are sma[illegible]