2

Correspondence Analysis, with Special Attention to the Analysis of Panel Data and Event History Data

Peter G. M. van der Heijden and Jan de Leeuw†*

We present correspondence analysis as an exploratory method that uses graphical representations to study the relation between rows and columns of a two-way table with non-negative entries. We present multiple correspondence analysis (MCA) as ordinary correspondence analysis of a so-called superindicator matrix. In this matrix, objects (e.g. persons) are in the rows, and each category of each variable has a separate column. MCA uses only the bivariate marginal frequencies to derive a representation for the columns. Therefore, it can handle data sets with many variables with many categories. We give special attention to panel data and event history data. We show how these types of data can be coded in three-way superindicator matrices with objects in the rows, categories of the variables in the columns, and time points in the layers.

This paper is a rewritten and adapted version of some chapters in van der Heijden's *Correspondence Analysis of Longitudinal Categorical Data* (D.S.W.O. Press, Leiden, 1987). For helpful comments, we are indebted to three anonymous reviewers, and we are grateful to Henriette Meerens for kindly providing us with her data. This research was partially supported by a grant from the Netherlands Organization for the Advancement of Pure Research (Z.W.O.) to the first author.

*University of Leiden
†University of California at Los Angeles

To analyze these data with MCA, we construct two-way codings. We discuss the implications of choosing specific two-way codings for the results of MCA. We compare Markov chain modeling of panel data with statistical methods for event history data. We show that MCA can easily handle data sets with many variables, categories, and time points and that it can be used to study individual differences, unlike most statistical approaches, in which persons are usually treated as replications. MCA is a very flexible exploratory tool of data analysis, and much research in this area is needed.

1. INTRODUCTION

In the last 15 years, there has been a growing interest in correspondence analysis (CA), a tool for the analysis of categorical data. CA has quite a long history (de Leeuw 1983). Since its development in 1935, it has been reinvented several times under different names. These approaches are formally identical, but their objectives, rationales, and procedures can be quite different. CA (also called *simple* CA) is intended for the analysis of two-way tables, and it is (formally) identical to reciprocal averaging, canonical correlation analysis, and simultaneous linear regressions (see, e.g., Nishisato 1980; Greenacre 1984). *Multiple* correspondence analysis (MCA) is a version of CA that is meant for the analysis of more than two variables; it is also known as homogeneity analysis and the quantification method (see also Tenenhaus and Young 1985). CA and MCA are considered the same in optimal scaling (Bock 1960) and in dual scaling (Nishisato 1980). The term *correspondence analysis* originated in France, where it is very popular. Compared with other approaches, CA places a heavy emphasis on geometrical representations, which is probably one of the reasons it became so popular, at least in France.

CA can handle many different types of data, such as paired-comparison, ranking, rating, and sorting data (see, e.g., Nishisato 1986). Data are coded into an appropriate two-way matrix (not necessarily a contingency table) so that CA of this matrix reveals some important aspects of the original data. The matrix is approximated in a least squares sense by a matrix of lower rank. This lower-rank approximation is studied in graphical

representations. Such representations show, among other things, the relation between the row and column entries of the matrix. In many applications the rows are individuals, and MCA gives a representation of *individual differences*; i.e., it shows how individuals differ in their relation to the columns. CA also finds *optimal quantifications* for the rows and columns of the matrix under study.

In this paper we will focus upon CA of longitudinal data. We speak of longitudinal data if one or more objects or phenomena are observed more than once using one or more variables. Our objective is to show how CA can be used to analyze longitudinal data and to compare it with more common approaches for analyzing these data. For *discrete* time, many models can be formulated as loglinear models (see Bishop, Fienberg, and Holland 1975; Plewis 1985). For *continuous* time, we find statistical models for the analysis of event histories (see Tuma and Hannan 1984; Allison 1984). In both the discrete and the continuous time approaches, the number of categories of the variable(s) under study may not be too large because of empty-cell problems. We will show that CA can easily deal with variables having a very large number of categories. We will concentrate on the analysis of panel data and event history data.

2. CORRESPONDENCE ANALYSIS OF CROSSTABLES

We will introduce correspondence analysis as a method for the analysis of crosstables. For details and proofs, we refer to the standard works of Nishisato (1980), Gifi (1981), Greenacre (1984), and Lebart, Morineau, and Warwick (1984). Consider the crosstable **P**, having values p_{ij}, where i ($i=1, \ldots, I$) indexes the I rows, j ($j=1, \ldots, J$) indexes the J columns, and $p_{ij} \geq 0$. We denote the margins as p_{i+} and p_{+j} for the rows and columns respectively (a "+" replaces an index when summed over the corresponding variable). In principle we can use CA to analyze any two-way matrix with non-negative entries. One type of crosstable for which CA is particularly suited is a contingency table. In this section we will assume that **P** is a table with proportions that add up to one: $p_{++}=1$. At the end of this section, we will discuss what tables can be analyzed with CA.

CA gives a graphical representation of **P**. In this representation, each row and each column of **P** has a separate point, and the configuration of points tells us what is going on in the data. To understand what is going on, we must know how distances between points relate to aspects in the data. The distance between two row points i and i' is a function of the differences between so-called *profiles* of these rows. The profile for row i is defined as the vector of conditional proportions p_{ij}/p_{i+}. The differences between the two profiles are weighted by $(1/p_{+j})$, thus bringing down the influence of the better-filled columns:

$$\delta^2(i,i') = \sum_j (1/p_{+j})\ (p_{ij}/p_{i+} - p_{i'j}/p_{i'+})^2. \qquad (1)$$

Distances $\delta(i,i')$ are called *chi-squared distances*. When $\delta(i,i')$ is large, the profiles of rows i and i' differ much; whereas when $\delta(i, i')$ is small, the profiles of row i and i' are similar, and we conclude that i and i' are related in the same way to the columns. The chi-squared distances between the rows show the *dependence* in matrix **P**: When the matrix is independent, all row profiles are identical and equal to the average row profile with values $p_{+j}/p_{++}=p_{+j}$, i.e., the marginal column proportions.

We want to find a representation of the row points in low-dimensional euclidean space. We can obtain a solution as follows. Without loss of generality, let $I>J$. First define $\mathbf{D}_r$ and $\mathbf{D}_c$ as diagonal matrices with marginal row and column proportions p_{i+} and p_{+j} respectively. Now $\mathbf{D}_r^{-1}\mathbf{P}$ is the matrix with *row* profiles. The I rows can be plotted as points in a J-dimensional *space* by using row vectors of $\mathbf{D}_r^{-1}\mathbf{P}$ as coordinates. By defining a weighted euclidean *metric* $\mathbf{D}_c^{-1}$ for this space, we get chi-squared distances between the rows (see Greenacre 1984 for more details). We can *center* this configuration of row points by subtracting from each row profile the average row profile with values p_{+j}. This can be done by using the matrix $\mathbf{D}_r^{-1}(\mathbf{P}-\mathbf{E})$ as coordinates, where **E** is an independent matrix: $e_{ij}=p_{i+}p_{+j}$. $\mathbf{D}_r^{-1}\mathbf{E}$ is a matrix with (identical) profiles of the column totals. By this centering, the rows span a $(J-1)$-dimensional subspace in the J-dimensional space, and the center of this space can be interpreted as the point for the *average row profile*.

Similarly, we can plot the J *columns* as points in an $(I-1)$-dimensional space with a weighted euclidean metric defined by $\mathbf{D}_r^{-1}$ by using $\mathbf{D}_c^{-1}(\mathbf{P}-\mathbf{E})^T$ as a matrix with coordinates. Since we assumed that $I>J$, and the columns are centered, the column points span a $(J-1)$-dimensional subspace of the I-dimensional space. The center can be interpreted as the point for the average column profile. In this way we can also study the dependence in $\mathbf{P}$ from the representation of the column points.

This study of the space of the rows and the columns is simplified greatly when a large part of the distances is displayed in a *low-dimensional* space. In CA this is accomplished as follows. First, weights are defined for the row and column points by p_{i+} and p_{+j} respectively. Then, new axes are defined in such a way that weighted squared projected distances to these axes are maximized for subsequent dimensions. If we denote the row coordinates in this new coordinate system as $\tilde{r}_{i\alpha}$ for row i on dimension α, $\lambda_1^2 = \Sigma_i p_{i+}\tilde{r}_{i1}^2$ is maximized for dimension 1, then $\lambda_2^2 = \Sigma_i p_{i+}\tilde{r}_{i2}^2$ is maximized for dimension 2, and so for further dimensions; the coordinates on distinct dimensions are uncorrelated. The same criterion is used for the columns. The weighting by p_{i+} is performed so that categories with large marginal proportions have more influence on the determination of the new axes. The new axes are found for the rows and columns simultaneously by performing a *generalized singular value decomposition* (see Greenacre 1984 for details). The generalized singular value decomposition is performed of

$$\mathbf{D}_r^{-1}(\mathbf{P}-\mathbf{E})\mathbf{D}_c^{-1} = \mathbf{R}\Lambda\mathbf{C}^T, \tag{2}$$

where Λ is a diagonal matrix with $(J-1)$ singular values λ_α in decreasing order; $\mathbf{R}^T\mathbf{D}_r\mathbf{R} = \mathbf{I} = \mathbf{C}^T\mathbf{D}_c\mathbf{C}$; $\mathbf{R}$ is of order $I\times(J-1)$ and has elements $r_{i\alpha}$ for category i on dimension α; $\mathbf{C}$ is of order $J\times(J-1)$ and has elements $c_{j\alpha}$ for category j on dimension α. Euclidean distances between the rows are chi-squared distances if scores $\tilde{\mathbf{R}} = \mathbf{R}\Lambda$ are used as coordinates, and euclidean distances between the columns are chi-squared distances if scores $\tilde{\mathbf{C}} = \mathbf{C}\Lambda$ are used as coordinates. By rewriting (2) using $\mathbf{R}^T\mathbf{D}_r\mathbf{R} = \mathbf{I} = \mathbf{C}^T\mathbf{D}_c\mathbf{C}$, we can find

$$\tilde{\mathbf{R}} = \mathbf{R}\Lambda = \mathbf{D}_r^{-1}\mathbf{P}\mathbf{C}, \tag{3a}$$

$$\tilde{\mathbf{C}} = \mathbf{C}\boldsymbol{\Lambda} = \mathbf{D}_c^{-1}\mathbf{P}^T\mathbf{R}, \tag{3b}$$

since $\mathbf{D}_r^{-1}\mathbf{E}\mathbf{C} = 0 = \mathbf{D}_c^{-1}\mathbf{E}^T\mathbf{R}$. Equations (3a) and (3b) are known as the *transition formulas*. Equation (3a) shows that row scores $\tilde{\mathbf{R}}$ can be derived from the column scores $\mathbf{C}$ by placing the row points in the weighted averages of the column points, where weights are defined by profile elements. Equation (3b) shows that column scores $\tilde{\mathbf{C}}$ can be derived from the row scores $\mathbf{R}$ by placing the column points in the weighted averages of the row points. Thus, *joint plots* of row and column points are sometimes constructed with coordinates $(\tilde{\mathbf{R}},\mathbf{C})$ or $(\mathbf{R},\tilde{\mathbf{C}})$. If the first choice is used, only the distances between row points in full-dimensional space are chi-squared distances; if the second choice is used, only the distances between column points are chi-squared distances. Another possibility is to use the pair $(\tilde{\mathbf{R}},\tilde{\mathbf{C}})$ as coordinates, so that both distances between rows and distances between columns are chi-squared distances. However, the disadvantage of this is that the singular values are used twice. This disadvantage is solved in the last possibility by using $(\mathbf{R}\boldsymbol{\Lambda}^{1/2}, \mathbf{C}\boldsymbol{\Lambda}^{1/2})$, but then neither the distances between the rows nor the distances between the columns are chi-squared distances.

The transition formulas (3a) and (3b) are a third way to understand how CA displays the dependence in $\mathbf{P}$: When for row i the profile value p_{ij}/p_{i+} is larger than the average profile value p_{+j}, row i "pulls" more than average on j, causing j to be placed nearer to i. By multiplying both values by p_{i+}, we find that i is nearer to j when $p_{ij} > p_{i+}p_{+j} = e_{ij}$, i.e., when the observed proportion for cell (i,j) is larger than the independent proportion. Independence can be seen to be *baseline model* in CA: When $\mathbf{P}$ is independent, all rows and columns fall into the origin, and the singular values are all zero.

Interpretation of geometrical CA solutions is often made easier by studying tables of *contributions* of dimensions and points. In these contributions, the squared singular values play a crucial role. First, it can be proved that the singular values relate in the following way to the Pearson chi-square statistic X^2 for testing independence:

$$\text{trace } \boldsymbol{\Lambda}^2 \; X^2/n, \tag{4}$$

where n is the sample size. This shows that CA splits the dependence in the matrix into a number of dimensions. The proportion displayed on dimension α is equal to $\lambda_\alpha^2/\Sigma_\alpha\lambda_\alpha^2$. Second, for each dimension α, $\lambda_\alpha^2 = \Sigma_i p_{i+}\tilde{r}_{i\alpha}^2 = \Sigma_j p_{+j}\tilde{c}_{j\alpha}^2$, so the squared singular value can for each dimension be split up over the rows and columns by studying proportions $p_{i+}\tilde{r}_{i\alpha}^2/\Sigma_i p_{i+}\tilde{r}_{i\alpha}^2$ and $p_{+j}\tilde{c}_{j\alpha}^2/\Sigma_j p_{+j}\tilde{c}_{j\alpha}^2$ respectively. These proportions add up to one for each dimension. A third type of contribution can be derived for a specific point by dividing its squared projected distance to the origin on a specific dimension by its total squared distance to the origin in full-dimensional space: Values $\tilde{r}_{i\alpha}^2/\Sigma_\alpha\tilde{r}_{i\alpha}^2$ show how good point i is represented on dimension α.

Another way to make interpretation easier is to use *supplementary information* (if available) on the rows or columns of the crosstable (see, e.g., Greenacre 1984). This supplementary information is fitted into the CA solution as a second step, after this solution has been derived from the crosstable. Consider a matrix $\mathbf{P}$ of order $I\times J$ and a supplementary matrix $\mathbf{S}$ of order $L\times J$ with extra information on the columns. Now, the L row profiles of $\mathbf{S}$, contained in $\mathbf{D}_r^{-1}\mathbf{S}$, are fitted into the solution with an equation similar to the transition formula (3a):

$$\tilde{\mathbf{R}}_s = \mathbf{D}_r^{-1}\mathbf{S}\tilde{\mathbf{C}}\mathbf{A}^{-1}, \tag{5a}$$

where $\tilde{\mathbf{R}}_s$ contains the coordinates of the L supplementary rows. The position of these supplementary points can give us further understanding of the configuration of the column points, from which they are derived. Similarly, we can derive coordinates for L^* supplementary columns using

$$\tilde{\mathbf{C}}_s = \mathbf{D}_c^{-1}\mathbf{S}^T\tilde{\mathbf{R}}\mathbf{A}^{-1}, \tag{5b}$$

where $\mathbf{D}_c^{-1}\mathbf{S}^T$ is a matrix of order $J\times L^*$ that contains the L^* profiles of supplementary columns.

So far, we have presented CA as a tool for making graphical representations of the dependence in a crosstable. It can be proved (see, e.g., de Leeuw 1973; Nishisato 1980; Greenacre 1984) that CA is formally identical to canonical analysis of contingency tables (see Kendall and Stuart 1967, ch. 33), an approach that emphasizes

the *quantification* of the categories of a two-way contingency table. For canonical analysis, we can rewrite (2) as

$$\mathbf{P} = \mathbf{E} + \mathbf{D}_r\mathbf{R}\Lambda\mathbf{C}^T\mathbf{D}_c = \mathbf{D}_r(\mathbf{1} + \mathbf{R}\Lambda\mathbf{C}^T)\mathbf{D}_c , \tag{6}$$

which is also known as the *reconstitution formula*. Now, the first column of **R** gives optimal scores for the row categories, the first column of **C** gives optimal scores for the column categories, and the first singular value λ_1 gives the *maximized correlation* between the optimally scored row and column variables. The second column of **R** (and **C**) gives optimal scores under the restriction that they are orthogonal to the first column of **R** (and **C**), and these optimal scores give the second maximized correlation λ_2, and so on for further dimensions.

There is a large interest in the relation between CA and other techniques for the analysis of categorical data. Goodman (1981) found that when the frequencies are derived from an underlying bivariate normal distribution (or a distribution that is bivariate normal after a suitable transformation of the rows and columns), the scores in the first column of **R** and **C** are approximately the same as the scores in the rows and columns of the log-multiplicative *RC* association model. Goodman (1985, 1986) discusses this in more detail and also introduces forms of CA as a model (see also Gilula and Haberman 1986). In our own work (van der Heijden and de Leeuw 1985), we have given more attention to the interpretation of (nonstatistical) CA as a tool for the *analysis of residuals* from independence, an aspect that is clearly seen in equation (6). Escofier (1984) extends CA by using the generalization

$$\mathbf{P} = \mathbf{Q} + \mathbf{S}_r\mathbf{R}\Lambda\mathbf{C}^T\mathbf{S}_c , \tag{7}$$

where **Q** is a matrix with estimates of expected proportions under some model that is less restrictive than independence, and $\mathbf{S}_r$ and $\mathbf{S}_c$ are diagonal matrices that do not necessarily consist of the margins of **P**. Thus, in van der Heijden and de Leeuw (1985), we use CA for the analysis of residuals from loglinear models in higher-way tables. We start by coding higher-way tables into two-way tables by stacking the categories of the original variables (i.e., when two of the original variables have 2 and 3 levels, the new row variable has $2\times3=6$ levels). It is easy to see that in this way the values of the matrix **E** are equal to estimates of expected frequencies

for the loglinear model in which the set of row variables are independent of the set of column variables. Instead of the independent matrix **E**, one could choose, for example, a conditional independence model and decompose residuals with (7). Other examples of this approach use quasi-independence, symmetry, and quasi-symmetry models. Overviews of this approach can be found in van der Heijden (1987) and van der Heijden, de Falguerolles, and de Leeuw (1989).

Goodman's approach to CA as a model and our residual analysis approach to CA are most useful when the table to be analyzed is a contingency table with frequencies derived under Poisson or (product-)multinomial sampling. But CA is also useful for many other types of tables, for example, the Burt table and the superindicator matrix, to be discussed in the next section. Another table will be analyzed as an example. Generally speaking, CA is a useful technique when the chi-squared distance is a useful measure for the (dis)similarity between the rows or between the columns of the matrix under study.

2.1. *Example*

For an example, we will analyze the matrix displayed in Table 1. It is a matrix of order $18\times(5\times5)$, having hours from 6:00 A.M. until midnight in the columns and the joint behavior of husbands and wives in the rows. In the cells, we find the total number of minutes in an hour that the husbands and wives in 326 couples spent at home (H), at work (W), travelling (T), shopping (S), and in other activities (O). The column totals all equal (326×60), because each of the 326 couples spends 60 minutes in an hour. The husband-wife pairs come from the National Travel Survey of 1980 conducted by the Dutch Central Bureau of Statistics (Moning 1983). From this survey, we selected the 326 couples in which both partners worked more than 25 hours a week and in which both kept diaries of their weekday activities. For more details, see van der Heijden (1987).

Table 1 tells us the specific hours in which activities were performed. We use CA to study this. In CA, each column profile is compared with the average column profile, having values 1/18. Each row profile is compared with the profile of the column

TABLE 1

The Number of Minutes in Each Hour that 326 Couples Spend in Various Combinations of Activities

	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	Totals
HH	17,345	11,013	4,226	3,027	2,261	2,191	3,414	3,279	2,704	2,556	3,896	9,378	13,548	13,124	12,233	13,173	14,407	16,399	148,174
HW	865	2,377	3,280	2,262	1,970	1,896	2,365	2,656	2,552	2,478	2,040	1,393	1,255	1,413	1,240	1,150	779	550	32,521
HS	0	0	0	0	70	10	25	0	9	0	5	137	10	4	0	0	0	0	270
HT	820	1,787	466	146	119	187	284	240	314	330	985	1,379	1,002	549	332	336	438	362	10,076
HO	10	39	90	60	20	0	60	25	35	65	284	348	506	650	898	826	625	325	4,866
WH	120	337	1,463	1,927	1,625	1,527	1,632	1,530	1,720	1,681	1,273	1,177	498	342	391	405	260	192	18,100
WW	60	1,118	6,773	10,547	11,178	11,354	8,186	8,474	9,483	8,740	5,279	1,272	486	560	575	335	130	120	84,670
WS	0	0	0	0	46	82	50	60	170	155	75	37	0	0	0	0	0	0	675
WT	20	585	704	400	326	393	556	621	287	409	1,221	451	31	70	77	30	40	0	6,221
WO	0	0	50	120	240	248	397	165	264	305	190	55	0	0	40	105	120	75	2,374
SH	0	0	0	25	222	231	305	140	190	198	164	320	120	184	80	60	30	0	2,269
SW	0	0	0	55	195	185	326	410	453	542	330	144	20	0	0	0	0	0	2,660
SS	0	0	0	35	30	15	85	60	52	85	221	155	5	20	95	30	0	0	888
ST	0	0	0	0	10	50	31	35	50	75	173	198	0	0	0	0	0	0	622
SO	0	0	0	0	0	0	0	5	0	30	0	25	0	50	0	0	0	0	110
TH	170	883	643	97	128	133	282	180	101	115	597	754	644	443	302	149	230	123	5,974
TW	0	416	1,156	429	330	300	828	829	482	708	963	431	175	108	65	55	26	0	7,301
TS	0	0	0	0	0	0	5	25	25	15	22	2	0	20	0	0	0	0	114
TT	150	960	629	175	160	163	193	395	203	352	866	1,050	352	360	357	149	185	247	6,946
TO	0	0	0	0	10	0	25	26	1	22	171	102	24	55	113	41	15	15	620
OH	0	30	30	130	65	0	43	50	55	43	150	174	252	685	930	540	410	249	3,836
OW	0	15	50	110	438	455	284	288	345	474	429	178	135	150	270	250	180	45	4,096
OS	0	0	0	0	0	0	9	0	20	0	6	20	10	42	30	0	0	0	137
OT	0	0	0	15	30	15	26	0	20	22	44	97	70	124	36	45	135	35	714
OO	0	0	0	0	75	125	149	67	25	160	176	283	417	607	1,496	1,881	1,550	823	7,834
Totals	19,560	19,560	19,560	19,560	19,548*	19,560	19,560	19,560	19,560	19,560	19,560	19,560	19,560	19,560	19,560	19,560	19,560	19,560	352,068

Note. Bivariate categories are in the rows, hours (starting with the hour from 6–7 A.M.) are in the columns. The categories are as follows: *H* = at home, *W* = at work, *T* = travelling, *S* = shopping, and *O* = other activities. Each joint activity is denoted by two letters, the first indicating the behavior of the wife, the second the behavior of the husband.

*12 minutes missing.

margins. Notice that this latter margin shows us that males generally spend more time working, whereas women spend more time shopping, etc. CA of Table 1 will not show this, since it concentrates on the departure from this margin.

CA gives eigenvalues $\lambda_1^2 = 0.47$ (76 percent of chi square), $\lambda_2^2 = 0.06$ (10 percent), and $\lambda_3^2 = 0.04$ (0.06 percent). To decide upon the number of dimensions that we are going to study, we can use the elbow or scree criterion, known from dimension-reduction techniques like factor analysis. It tells us that we can restrict attention to the first dimension only. Results are summarized in Figure 1. The hours are presented horizontally and the quantifications on the first dimension are presented vertically. The activities are represented on the vertical line on the left, and the hours are represented by points that are connected for adjacent hours. The morning hours 6, 7, and 8 (i.e., from 6:00 A.M. until 9:00 A.M.) and the evening hours 18 to 23 are quantified negatively; the other hours are quantified positively. Thus, in the morning and evening, the couples perform activities different from the activities they perform in the hours between.

The quantifications of the row categories show what these activities are. We use the transition formulas (3a) and (3b). To interpret the row categories, it is helpful to study the contributions of these points to the first dimension. This shows contributions of 0.38 for *HH* (both partners at home) and 0.43 for *WW* (both partners at work). Thus, more than 80 percent of the chi square displayed on this dimension stems from the fact that rows *HH* and *WW* depart from the average, being the profile with values 1/18. We now understand the quantification of the hour points: In the morning and in the evening, husbands and wives are both at home more than the average for the whole day, whereas during working hours, they are both at work more than the average for the whole day. The time-point quantifications, which all have about equal contributions, are determined mainly by these two points. There is a dip during lunch time because some persons go home then. Other states do not contribute much. Only the state in which both partners do other things (*OO*) takes account of another 0.04 and is quantified very negatively, since these activities are performed more than average in the evening. This does not mean that we should not interpret the position of the other states; they do not contribute

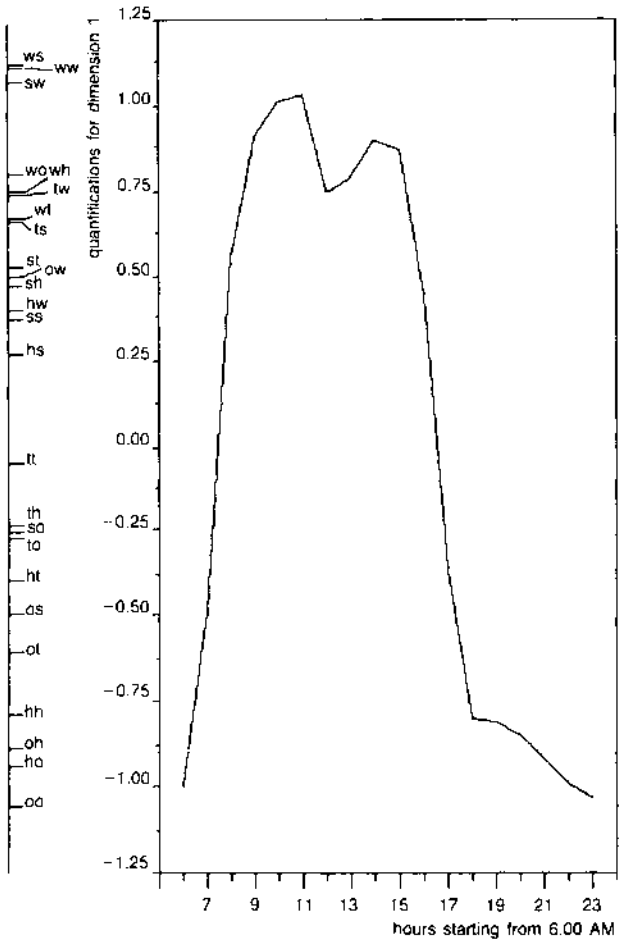

FIGURE 1 Analysis of data in Table 1. Quantifications on first dimension for the rows and columns. The *J* activity quantifications are displayed on the vertical line on the left; the *P* period quantifications are connected by a solid line on the right.

much because they have a lower mass than the categories *HH*, *WW*, and *OO*. For example, the state in which one partner shops while the other works (*WS* and *SW*) occurs more than average during working hours, whereas shopping together (*SS*) occurs more than average after working hours but before the evening (in the Netherlands most shops close at 6:00 P.M.). Furthermore, discrepancies between "opposed" states such as *HT* and *TH* can be

studied: This shows, for example, that in the morning, women stay at home somewhat longer than men and that in the evening, they are at home again earlier. These latter interpretations are more speculative, because these points have little influence on the solution, and it is possible that higher dimensions lead us to a different interpretation. Therefore, it is generally advisable to check conclusions like these in the data or to study whether these points have large contributions on higher dimensions. CA is used here to obtain only a global idea of the most important departure from independence (76 percent is displayed). Higher dimensions show mainly unstructured peculiarities in the departure.

Note that it would have been unusual to study Table 1 with loglinear analysis, because the assumption of independent observations (minutes) is clearly violated. If the observations in Table 1 were independent, CA could be interpreted as giving a representation of the residuals from the loglinear model in which hours are independent of the joint behavior of the couples (see above, and van der Heijden and de Leeuw 1985).

3. MULTIPLE CORRESPONDENCE ANALYSIS

There are many ways to introduce multiple correspondence analysis (MCA). Standard references are Benzecri (1973), Nishisato (1980), Gifi (1981), Greenacre (1984), and Lebart et al. (1984). Here we will discuss three ways. First, we will introduce MCA as CA of a so-called indicator matrix. In this context we will emphasize a chi-squared distance interpretation of MCA. Second, we will show that MCA can be interpreted as principal components analysis (PCA) of categorical data. Here the quantification interpretation of MCA will be stressed. Third, we will introduce MCA as CA of a so-called Burt matrix. This will make it easier to relate MCA to the decomposition of chi square and to loglinear analysis.

3.1. *Correspondence Analysis of a Superindicator Matrix*

Data are very often coded into a matrix with objects as rows and variables as columns. The objects can be, for example, persons, schools, countries, or household units. The variables denote aspects on which the objects are measured. We will assume that the

TABLE 2
A Small Example of a Categorical Data Matrix (Panel A) and its Superindicator Matrix (Panel B)

Panel A			Panel B								
			a	*b*	*c*	*p*	*q*	*r*	*u*	*v*	*w*
a	*q*	*w*	1	0	0	0	1	0	0	0	1
a	*r*	*w*	1	0	0	0	0	1	0	0	1
c	*r*	*v*	0	0	1	0	1	0	0	1	0
b	*q*	*u*	0	1	0	0	1	0	1	0	0
a	*r*	*w*	1	0	0	0	0	1	0	0	1
b	*p*	*u*	0	1	0	1	0	0	1	0	0
c	*r*	*w*	0	0	1	0	0	1	0	0	1
c	*p*	*v*	0	0	1	1	0	0	0	1	0
c	*q*	*w*	0	0	1	0	1	0	0	0	1
a	*p*	*v*	1	0	0	1	0	0	0	1	0

Note. There are ten objects (rows), and three variables, each with three categories.

resulting measures are categorical. In Table 2, panel A, we present an example of a data matrix for ten persons and three variables. The entries of the matrix, $\mathbf{X}$, are simply letters: i.e., a, b, c for variable 1, p, q, r for variable 2, and u, v, w for variable 3. The matrix has columns $\mathbf{X}^m$ ($m=1, \ldots, M$) and elements x_{im}, where i ($i=1, \ldots, I$) indexes the n rows. We can transform each column (variable) $\mathbf{X}^m$ into a so-called *indicator matrix* $\mathbf{G}^m$ with elements g_{ij}^m, where $g_{ij}^m = 1$ if object i falls into category j ($j=1, \ldots, J$) of variable m, and $g_{ij}^m = 0$ otherwise. Variable m has k_m categories. By concatenating the M indicator matrices $\mathbf{G}^m$ horizontally, we get a so-called *superindicator matrix*, denoted by $\mathbf{G}$. $\mathbf{G}$ has n rows and $k = \Sigma_m k_m$ columns. In our example, the matrix $\mathbf{G}$ has nine columns (see Table 2, panel B). CA of $\mathbf{G}$ is known as MCA. We will now apply this by discussing the properties of CA of $\mathbf{G}$.

CA gives a graphical representation of both row profiles and column profiles of a matrix. So the *n row profiles* in $\mathbf{D}_r^{-1}\mathbf{G}$ are plotted as points in a k-dimensional *space*. Notice that because each row in $\mathbf{G}$ has m 1s, $\mathbf{D}_r = \mathbf{I}/m$, and a row profile is a vector with m values equal to $1/m$, the other values being 0. Also, all row profiles

have identical weights. The *metric* of this space is defined by $\mathbf{D}_c^{-1}$, where $\mathbf{D}_c$ has diagonal elements $(\Sigma_i g_{ij}^m)/nm$, $\Sigma_i g_{ij}^m$ being the marginal frequency of category j of variable m, and nm being the total number of 1s in the superindicator matrix. Two row points will be near each other when they have similar profiles, i.e., when the objects fall into the same categories of the m variables. The *column* profiles of **G** are plotted as points in n-dimensional space. Associated with these points are masses $(\Sigma_i g_{ij}^m)/\text{nm}$ that are proportional to the marginal frequencies of the categories. Two column profiles will be near each other when there are many objects that fall either into both of these categories or into neither of these categories. The *dimensionality* of the subspaces that are spanned by the row points and the column points is $\min((n-1), (k-m))$, where $(n-1)$ is the dimensionality of the centered n-dimensional column space and $(k-m)$ is the dimensionality of the centered row space, since for each variable m, $\Sigma_j g_{ij}^m = 1$, and therefore centering eliminates m dimensions. Related to this is a special property of the category scores. For each variable on each dimension, they add up to zero when we weight them with the marginal frequencies of the categories: $\Sigma_j g_{ij}^m c_{j\alpha}^m = 0$ for each dimension α and each variable m, where $c_{j\alpha}^m$ is the quantification of category j of variable m on dimension α.

Because **G** is a binary matrix, the relation between the row points and column points simplifies. Equation (3b) now shows that if the normalization $(\mathbf{R},\tilde{\mathbf{C}})$ is chosen, a column point, i.e. a category, is in the *centroid* of the objects that fall into that category. Equation (3a) shows that if $(\tilde{\mathbf{R}},\mathbf{C})$ is chosen, an object point is in the average of its categories. In most applications, normalization $(\mathbf{R},\bar{\mathbf{C}})$ is chosen.

In most applications, *supplementary information* on the objects (e.g., sex, place of birth) is also available. When the supplementary information is (made) categorical, we can code it into a (super)indicator matrix **S** and apply equation (5b). Thus, we find supplementary category points that are in the average of all objects that fall into these categories (see Greenacre 1984).

3.2. *Principal Components Analysis of Categorical Data*

One way to define PCA is as follows. Consider a set of m quantitative variables collected in a matrix **Q** having columns $\mathbf{Q}^m$.

Then, the first principal component, $\mathbf{R}^1$, is the linear combination of the m variables $\mathbf{Q}^m$ that maximizes the average of the squared correlations $\lambda_1^2 = \Sigma_m (\text{cor}(\mathbf{R}^1,\mathbf{Q}^m))^2/m$. The average λ_1^2 is called the first eigenvalue, and the correlations $\text{cor}(\mathbf{R}^1,\mathbf{Q}^m)$ are known as component loadings. The second principal component, $\mathbf{R}^2$, is the linear combination that maximizes $\lambda_2^2 = \Sigma_m(\text{cor}(\mathbf{R}^2,\mathbf{Q}^m))^2/m$ under the restriction that it is uncorrelated with the first, and so on for further dimensions.

MCA can easily be fitted into this framework. Consider the matrix $\mathbf{X}$, with categorical measures in the cells (see Table 2, panel A, for an example). Suppose we have analyzed the table $\mathbf{G}$, derived from $\mathbf{X}$, with CA. This gives us scores collected in $\mathbf{R}$ and $\mathbf{C}$ and eigenvalues collected in $\mathbf{\Lambda}^2$. Consider now only the first dimension of CA, i.e., the first column of row scores $\mathbf{R}^1$ and column scores $\mathbf{C}^1$, where the score of category j of variable m is denoted with c_{j1}^m. We can use the category scores c_{j1}^m to derive from the original matrix $\mathbf{X}$ a *quantified* data matrix $\mathbf{Q}$ by replacing the categories in $\mathbf{X}$ by their corresponding quantifications in $\mathbf{C}^1$. The relation between PCA and MCA is such that these category scores are optimal in the sense that $\lambda_1^2 = \Sigma_m(\text{cor}(\mathbf{R}^1,\mathbf{Q}^m))^2/m$ is maximized. Using $\mathbf{C}^2$, we can construct a newly quantified data matrix $\mathbf{Q}$ by filling in these category scores, and these category scores are optimal in the sense that $\lambda_2^2 = \Sigma_m(\text{cor}(\mathbf{R}^2,\mathbf{Q}^m))^2/m$ is maximized under the restriction that $\mathbf{R}^1$ is uncorrelated with $\mathbf{R}^2$, and so on for further dimensions. This shows that MCA is a PCA for categorical variables that are optimally quantified for each subsequent principal component.

In the description of ordinary CA, we discussed three types of *contributions* that could be computed to simplify the interpretation of the CA solution. The first type was the proportion of chi square that was decomposed in each dimension. This type of contribution is less useful in MCA, since the last dimensions of the MCA solution can be shown to be artifacts due to linear dependencies (see, e.g., Greenacre 1984, ch. 5). Therefore, distances on these last dimensions should not be interpreted. The second type was the decomposition into subsequent dimensions of the distance of a point to the origin in full-dimensional space. Now that we know that the last dimensions in the MCA solution are artifacts, this type of contribution is not very useful either. Only the third type of contribution is useful: the contribution of individual row and column

points to a dimension. For the rows of **G**, this contribution is easily derived, because all row weights, defined by $\mathbf{D}_r$, are identical. For the columns of **G**, it is most useful to sum these contributions per variable so that it is clear how much a particular variable (instead of a particular category) contributes to a specific dimension. The contribution $(g^m_{+j}/nm)(\tilde{c}^m_{j\alpha})^2/\lambda^2_\alpha$ shows the contribution of category j of variable m to dimension α. By adding up over the categories j in variable m, we have the contribution of variable m: $\Sigma_j(g^m_{+j}/nm)(\tilde{c}^m_{j\alpha})^2/\lambda^2_\alpha$. This quantity is closely related to the squared correlation between the quantified variable m and the row scores for dimension α, since $(\text{cor}(\mathbf{R}^\alpha,\mathbf{Q}^m))^2 = \Sigma_j(g^m_{+j}/n)(\tilde{c}^m_{j\alpha})^2$ (for a proof, see Tenenhaus and Young 1985).

3.3. *Correspondence Analysis of the Burt Matrix*

Further insight into MCA can be obtained when we consider it as CA of the so-called Burt matrix $\mathbf{B}=\mathbf{G}^T\mathbf{G}$. Table 3 displays the Burt matrix for our little example. The Burt matrix can be considered a concatenation of all *univariate* margins (as diagonal submatrices on the diagonal of the Burt matrix) and *bivariate* margins. CA of the matrix **G** is related to CA of the matrix $\mathbf{B}=\mathbf{G}^T\mathbf{G}$ in the following way: Column scores **C** derived from analysis of **G** are equal to column (and row) scores **C** derived from analysis of **B**. The only difference is in the singular values: For these, we have $\lambda_{\alpha(\mathbf{B})} = \lambda^2_{\alpha(\mathbf{G})}$, where $\lambda_{\alpha(\mathbf{B})}$ is the singular value for the analysis of

TABLE 3
The Burt Matrix for the Example in Table 2

	a	*b*	*c*	*p*	*q*	*r*	*u*	*v*	*w*
a	4	0	0	1	1	2	0	1	3
b	0	2	0	1	1	0	2	0	0
c	0	0	4	1	1	2	0	2	2
p	1	1	1	3	0	0	1	2	0
q	1	1	1	0	3	0	1	0	2
r	2	0	2	0	0	4	0	1	3
u	0	2	0	1	1	0	2	0	0
v	1	0	2	2	0	1	0	3	0
w	3	0	2	0	2	3	0	0	5

the Burt matrix, and $\lambda_{\alpha(\mathbf{G})}$ is the singular value for the analysis of the superindicator matrix $\mathbf{G}$.

Ordinary CA decomposes the departure from independence. When we apply this to the Burt matrix, we see that an "independent" matrix derived from the Burt matrix is in fact a matrix with marginal independence for all pairs of variables. So MCA studies the departure from bivariate marginal *independence*, restricting attention to the bivariate marginal *dependence*. Now, the relation with chi square displayed in equation (4) needs to be changed: This relation becomes $\Sigma_\alpha \lambda_{\alpha(\mathbf{B})} = \Sigma_\alpha \lambda^2_{\alpha(\mathbf{G})} = (k-m)+\Sigma_m \Sigma_{m'} X^2_{mm'}/n$, where $X^2_{mm'}$ is the chi square for testing independence in the marginal table of variables m and m'. It also shows, in loglinear analysis terminology, that second- and higher-order interactions are ignored. There are circumstances in which this is a serious drawback. However, by restricting attention to marginal bivariate dependence, we can study the relation between many variables with many categories without running into empty-cell problems. Therefore, MCA has a large range of possible applications.

The Burt matrix also gives us a further understanding of the *quantification interpretation* of MCA. In CA the categories of the two variables are quantified in such a way that the correlations are maximized. By quantifying, we can derive for each dimension a correlation for the contingency table. In MCA the variables are all quantified optimally under the restriction that the quantifications of the categories of a variable are identical for the relations with all other variables. With these quantifications, a correlation coefficient can be derived from each subtable of the Burt matrix, and this matrix can be reduced to a correlation matrix for each dimension.

3.4. *Missing Data*

So far, we have discussed MCA with no missing values. When there are missing values, it is not known in which category an object falls. Most often this is remedied by coding 0 in the indicator matrix for all categories of the missing variable. In many instances, formulas will change a little, because with missing values, row margins are not the same for all rows. MCA with missing

values is treated in detail by Meulman (1982; see also Nishisato 1980, ch. 8; Greenacre 1984, ch. 5).

4. THREE-WAY SUPERINDICATOR MATRICES

In the sequel, we concentrate on the analysis of longitudinal data. We speak of longitudinal data when a group of objects is observed more than once at the same time points on the same (group of) variable(s). We assume that the measures are categorical. Therefore, longitudinal data can be coded into a *three-way superindicator matrix* **G**, having elements g_{ijt}^{m}, where $g_{ijt}^{m} = 1$ if object i falls into category j of variable m at time point t ($t = 1, \ldots, T$), and $g_{ijt}^{m} = 0$ if not. The three modes of **G** are the object mode indexed by i, the category mode indexed by a combination of j and m, and the time mode indexed by t, so only the last mode is new. This way of coding longitudinal data is not very restrictive. The only restrictive feature is that in principle, all objects must be measured at the same (and not at different) time points. If they are not, we have to code some objects at specific time points as missing, or use an extra category for them. However, in both possibilities, results can be influenced to some extent by the fact that a number of objects are missing at specific time points.

The coding is useful not only in *discrete* time applications. Observations in *continuous* time can be coded into a three-way superindicator matrix when we keep in mind that in practice, the precision with which time is observed is limited: Continuous time observations are measured in days, hours, minutes, or seconds. Large precision implies only that the number of time points is very large, making the time mode of the matrix **G** very large. Three-way superindicator matrices were first introduced by Saporta (1981); see also de Leeuw, van der Heijden, and Kreft (1985) and van der Heijden (1987).

Our approach to the analysis of **G** is to recode the three-way data in **G** into a *two-way matrix* and to analyze this two-way coding with MCA. In principle there are three ways to construct two-way codings: First, we can concentrate on separate slices of the three-way data block. Second, we can concatenate these slices. Third, we can concentrate on margins of the three-way matrix **G**. In this

paper we will not analyze slices but concentrate instead on concatenations of slices and on marginal tables.

There are three ways to *concatenate slices*. First, we can concatenate the T time-point slices horizontally so that we obtain a broad matrix with objects in the rows and categories of variables at specific time points in the columns. We will refer to this matrix as the *BROAD* matrix and denote it as $\mathbf{G}^{i(jt)}$, having elements $g^m_{i(jt)}$. Second, we can concatenate the T time-point slices vertically, yielding a *LONG* matrix with categories of variables in the columns and object-time-point combinations in the rows. The *LONG* matrix will be denoted as $\mathbf{G}^{(it)j}$, with elements $g^m_{(it)j}$. Third, we can construct a matrix with time points in the columns and combinations of objects and categories of variables in the rows. However, relations between the rows and columns of this matrix are not easily interpretable; therefore, we will skip this possibility.

For the *marginal tables*, we also have three possibilities. First, we can add up over the time points. This gives a matrix denoted as $\mathbf{G}^{IJ}$ with objects in the rows and categories of variables in the columns. The entries g^m_{ij+} of $\mathbf{G}^{IJ}$ are frequencies indicating the number of times that object i falls into category j of variable m. Second, we can add up over the objects so that we have a matrix with time points in the rows and categories of variables in the columns. This matrix will be denoted as $\mathbf{G}^{TJ}$, and entries g^m_{+jt} are frequencies indicating the number of objects that fall into a specific category at a specific time point. The third marginal matrix, obtained by adding up over the category mode, is the uninteresting matrix with all entries equal to $g^+_{i+t} = m$, the number of variables.

So we conclude that the three-way matrix $\mathbf{G}$ gives us four interesting two-way matrices, namely, the *BROAD* and the *LONG* matrices obtained by concatenating slices, and the marginal matrices $\mathbf{G}^{IJ}$ and $\mathbf{G}^{TJ}$. Below, we will discuss MCA of panel data and of event history data. Specifically, we will discuss the aspects of the data that are revealed when we analyze each of the four possible two-way matrices and compare this to more usual techniques for the analysis of these data.

5. MULTIPLE CORRESPONDENCE ANALYSIS OF PANEL DATA

We speak of panel data if a group of objects is measured more than once at specific time points using one or more variables. Such data can always be coded into the three-way superindicator matrix **G**. We will first consider univariate panel data. Then we will discuss the analysis of multivariate panel data.

5.1. *Univariate Panel Data*

We can consider panel data as univariate when either the number of variables is one or a composite variable is constructed by stacking the categories of two or more variables into one new variable. For the general case, we saw that the three-way superindicator matrix **G** could be reduced to four interesting two-way codings of the data. For this specific univariate application, the *LONG* matrix $\mathbf{G}^{(it)j}$ gives a trivial solution with all singular values equal to one, since $\mathbf{G}^{(it)j}$ has rows with one 1, the other values being zero.

The most interesting analysis is the analysis of *BROAD matrix* $\mathbf{G}^{i(jt)}$. This matrix has one row for each object and one column for each category at each time point. Each row has T values of one, the others being zero. The chi-square distance interpretation of MCA indicates that each object (row) is compared with the average row profile, being the profile of the column margins g_{+jt} (we skip the superscript m, since there is only one variable). These column margins specify the distributions of objects over the categories at the T time points. The profile of these column margins is placed in the origin. Note that we do not study the structure within the column margins, only the departure from this average. Objects are placed near each other when their profiles depart in the same way from the average profile, and they are placed far from each other when they depart in different ways. A separate analysis of the column margins of $\mathbf{G}^{i(jt)}$ can be useful to complement the analysis of $\mathbf{G}^{i(jt)}$ itself. In this way we can understand from what the objects depart. The column margins are collected in the *marginal matrix* $\mathbf{G}^{TJ}$. This matrix shows how categories relate to time points. CA of $\mathbf{G}^{TJ}$ compares the T distributions of objects with the average

distribution over all time points given by the column margins of $\mathbf{G}^{TJ}$.

When we interpret the MCA of $\mathbf{G}^{I(JT)}$ as a *PCA of nominal variables*, the variable that is measured at T time points is considered in $\mathbf{G}^{I(JT)}$ as a set of T distinct variables that will simultaneously receive an optimal quantification. So MCA gives T distinct sets of quantifications of this variable, one set for each time point t. This interpretation of MCA of $\mathbf{G}^{I(JT)}$ is closely related to MCA of the marginal matrix $\mathbf{G}^{IJ}$. Gifi (1981, ch. 10) shows that *equality constraints* (i.e., under which two or more rows/columns receive identical quantifications) can be obtained with ordinary CA programs by simply replacing the two or more rows/columns by one row/column that is the sum of the original rows/columns. Therefore, the analysis of the *marginal matrix* $\mathbf{G}^{IJ}$ gives a solution for the analysis of $\mathbf{G}^{I(JT)}$ in which category j obtains identical quantifications for each time point t. The transition formula (3a) shows that if in the analysis of $\mathbf{G}^{I(JT)}$ the category quantifications of corresponding categories do not differ much, the row scores $\mathbf{R}$ (i.e., the scores for the objects) for the analysis of $\mathbf{G}^{I(JT)}$ will be approximately equal to those for $\mathbf{G}^{IJ}$.

We conclude that for the three-way superindicator matrix, there are three interesting two-way matrices to be analyzed. The analysis of $\mathbf{G}^{I(JT)}$ is in general most interesting, because it shows how objects depart in different ways from the average. This analysis can be supplemented by the analysis of $\mathbf{G}^{TJ}$, giving more insight into this average. The analysis of $\mathbf{G}^{IJ}$ shows us a constrained analysis of $\mathbf{G}^{I(JT)}$.

So far, we have only considered matrices that can be derived from the three-way matrix $\mathbf{G}$ by adding up or concatenating slices of the data block. For the special case in which the number of time points $T=2$, we can also construct a *transition matrix* with the categories for $t=1$ in the rows and the categories for $t=2$ in the columns. It can be proven that an ordinary CA of such a transition matrix comes to the same as MCA of the matrix $\mathbf{G}^{I(JT)}$, in the sense that the former solution can be derived from the latter (see, e.g., Gifi 1981; Greenacre 1984). This property holds for contingency tables in general. We can also use CA to analyze panel data with three or more waves by creating a contingency table with a new composite variable, with the stacked categories of the first few time

points in the rows and the categories of the last time point in the columns. For an example, and its relation to Markov chain models, see van der Heijden (1987). We will now discuss an MCA example for three time points and then discuss the relation of MCA of univariate panel data to Markov chain models.

5.2. *Example*

In 1987 and 1988, Meerens, Boer, and Tan (1988) conducted a panel study of the decrease in income resulting from unemployment or medical disability. The first interview was conducted in January 1987, the second in July 1987, and the third in January 1988. There were 402 respondents at all three time points. Longitudinal information was collected for many variables (for more details, see Meerens et al. 1988). We concentrate here on the question, "Consider a person receiving money from social security. How much may such a person *according to your standards* earn in a month without notifying the social security office?" (In the Netherlands, earning extra money without notifying the social security office is illegal.) At time 1, the possible answers are "no idea," "nothing," "less than f100," "between f100 and f249," "between f250 and f499," "more than f500," and "unlimited, or could not specify an amount." (One Dutch guilder, denoted as f1, equals roughly \$0.5.) Because of the divers answers falling into this last alternative, at time points 2 and 3 two extra alternatives were added to this list: "the same amount as when I worked" and "I know that it is allowed, but I don't know how much." So there are $7\times9\times9=567$ possible response patterns, only 204 of which were used. For this reason, more traditional approaches that use the three-way matrix of order $7\times9\times9$ cannot be applied. The sequence of analyses below is in our opinion typical for an adequate study of panel data with MCA.

We start by coding the 402 profiles into a three-way indicator matrix of order $402\times9\times3$, of which at time 1 two categories are empty by design. The first analysis we perform is the analysis of the marginal matrix $\mathbf{G}^{TJ}$ of order 9×3. This matrix is displayed in Table 4. Although this matrix can easily be studied by eye only, we will do a CA for expository purposes. CA gives a solution that is dominated by a distinction between time 1 and times 2 and 3:

TABLE 4
Marginal Frequencies of Responses at Times 1, 2, and 3 to the Question About Social Security Income

Response	Time 1	Time 2	Time 3
No idea	50	6	13
Nothing	61	66	56
Less than f100	19	19	21
Between f100 and f249	64	68	56
Between f250 and f500	110	92	87
More than f500	77	45	55
Unlimited	21	13	16
Same amount as when I worked	—	68	57
I know it is allowed, but I don't know how much	—	25	41
Totals	402	402	402

This is due to the two categories that are zero by design at time 1 and to the "no idea" category, which has a high frequency at time 1 and a low frequency at times 2 and 3. So this analysis shows a somewhat trivial result. In de Leeuw and van der Heijden (1988), we discuss a procedure for the analysis of incomplete tables that is closely related to the loglinear quasi-independence model. Practically, it implies that independent values are imputed into the cells that make the table incomplete. For our example, these independent values are 81.97 and 43.28. When we analyze the table with the imputed values, we find a dominant eigenvalue of 0.13 (93 percent). The time points are quantified as −0.50, 0.49, and 0.16 for time 1 to time 3. Now the main difference is between the profiles for times 1 and 2, time 3 being in between. Categories "no idea," "greater than f500," and "unlimited" receive negative quantifications, showing that frequencies of these categories are particularly large at time 1 and become smaller at time 2 (and to a lesser extent, at time 3). On the other hand, the categories "nothing," "less than f100," and "f100–f249" are quantified positively, showing that the number of persons that fell into these categories at time 2 (and to a lesser extent, at time 3) became larger.

Now that we have some insight into the changes at the group level, we can study individual differences in changes with MCA. We do this by analyzing the *BROAD* matrix of order 402 × 25 (we can omit the two empty columns). MCA gives first few eigenvalues 0.621, 0.557, 0.514, 0.467, 0.441, and 0.411. We study only two dimensions here, thus using MCA only to show the most important information in the data. The first two dimensions are shown for the persons (rows) in Figure 2 and for the categories (columns) in Figure 3. These two figures are related in the sense that each category point is in the average of the persons that fall into it. In Figure 2, each different response pattern is represented by a point, so there are 204 distinct points. For this data set, the points are placed in a triangular configuration, and most points are located in the left corner. The chi-squared distance interpretation implies that response patterns (i.e. persons) are placed closer together when they have many elements in common (here, they are identical when

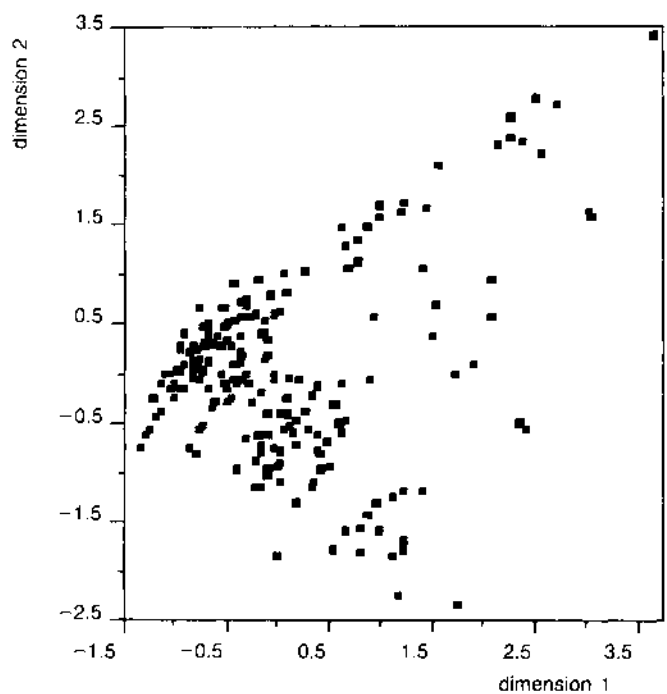

FIGURE 2 Row points (for objects/response patterns) in two dimensions.

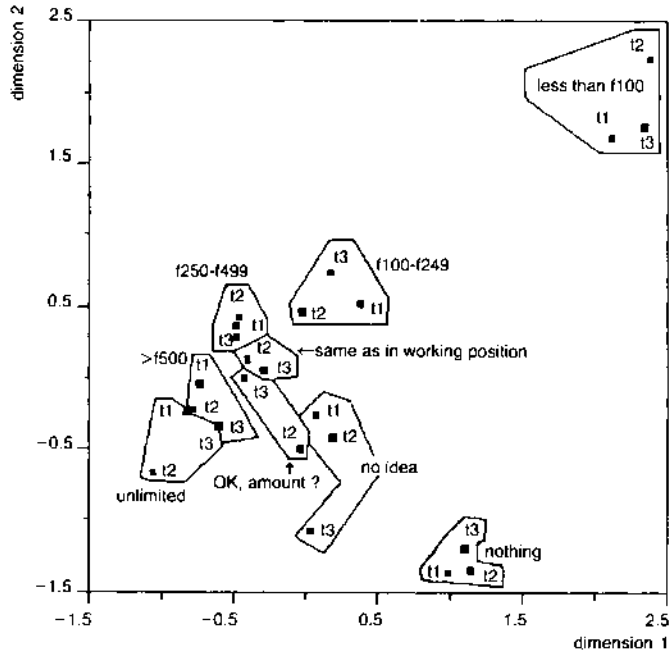

FIGURE 3 Category points in two dimensions.

they have three elements in common), and they are placed farther apart when they have little in common. We can interpret the form of the configuration of points when we study the plot of category points. Here we see in the upper right corner the categories for "less than f100," in the lower right corner "nothing," and in the left corner the other categories. Category points are close together when they have many persons in common; therefore, for example, the categories "less than f100" at times 1, 2, and 3 have many persons in common. This shows that persons who fall into "less than f100" at one of the time points very often fall into this category at other time points. When these persons change categories, they are most likely to change to nearby categories, since these categories have the most persons in common with the "less than f100" category. For the persons falling at least once in "less than f100," these are the categories for "f100–f249." Persons falling at least once in "less than f100" will seldom change to other categories, because "less than f100" is relatively far from all other categories. The configuration of the remaining points can be interpreted in a similar

fashion: A curve can be drawn from "less than f100" via "f100–f249," "f250–f499," "f500 or more," "unlimited," "no idea" to "nothing." Figure 2 shows that most persons have a response pattern with categories that are in adjacent clusters in this curve. However, there are several exceptions: There are also points between the upper right and lower right corners representing persons who change from "nothing" to "less than f100." That the distance between "less than f100" and "f100–f249" is much larger than the distance between "f100–f249" and "f250–f499" indicates that more changes occur between categories of the latter pair than between those of the former pair. Most changes occur between the categories in the left corner. The new categories "the same amount as when I worked" and "I know that it is allowed, but I don't know how much" are between all other categories, indicating that they are found together with all other categories.

For all categories, the points for times 1, 2, and 3 are near each other, indicating that relatively many persons remain in the same state. It is dangerous to interpret the small differences between the locations of a category at times 1, 2, and 3 because we are looking at a two-dimensional representation of a high-dimensional configuration, and although these two dimensions are most important, much information is hidden in higher dimensions. However, on both dimensions it is generally true that the points for time 2 are further away from the center than the points for times 1 and 3. This is also shown by the contributions of the time points to the dimensions: For times 1, 2, and 3 respectively, these are 0.317, 0.358, and 0.326 for dimension 1 and 0.307, 0.388, and 0.305 for dimension 2. This shows that the separation of person points is due more to their scores at time 2 than to their scores at times 1 and 3 (see the next section on Markov chain models and multiple CA). These differences are very small. This is shown by the fact that the analysis of $\mathbf{G}^{\mu}$ (a constrained analysis of the *BROAD* matrix) gives almost the same configuration of person points: For the first dimension, the correlation between the person scores for the constrained and the unconstrained analyses is 0.996; for the second dimension, it is 0.989.

As a second step in the multiple CA of the *BROAD* matrix, we will relate supplementary information to the solution: the age of the persons (in four categories), their sex, and whether they

stopped working for medical reasons (M) or became unemployed (U). We relate this information to the person points to find out whether persons with specific characteristics can be found in a specific part of the configuration. Thus, we use the supplementary information to explain differences between persons' scores on the longitudinal variable. We have derived from the three supplementary variables one new composite variable having $4\times2\times2=16$ categories, and we have computed averages for each of these 16 categories. (This strategy can only be applied when the product of all levels of the supplementary variables is not too large. If the product is large, we have to compute averages for each level of each variable separately.) See Figure 4. The categories discriminate most in the direction from the left corner to the lower right corner. In general, medically unfit persons more often state "nothing," whereas unemployed persons more often mention high amounts. There doesn't seem to be much difference by sex: Corresponding male and female points are quite near each other. There also seems to be a tendency for some unemployed groups to lie near the upper right corner (i.e., a small amount is allowed).

5.3. *Multiple Correspondence Analysis and Markov Chain Models*

In the presentation of MCA as a CA of the Burt matrix, we explained that MCA concentrates on the bivariate marginal dependence in the higher-way contingency table. In terms of loglinear analysis, a Burt matrix gives an adequate summary of the higher-way contingency table if a model with only first-order interactions fits the data reasonably well. If the bivariate margins are not sufficient for an adequate description of the data, then some of the information in the higher-way table is ignored.

Let's now consider Markov chain models. For discrete time, Markov chain models can be formulated in terms of loglinear models (see Bishop et al. 1975; Plewis 1985) and hence in terms of fitted margins. Assume that we have three-wave panel data coded into a three-way contingency table with elements f_{ijk}. If the data in the three-way matrix can be described by a *nonstationary* first-order Markov chain, then the two marginal tables with elements f_{ij+} and f_{+jk} are sufficient to describe the data (cf. Plewis 1985).

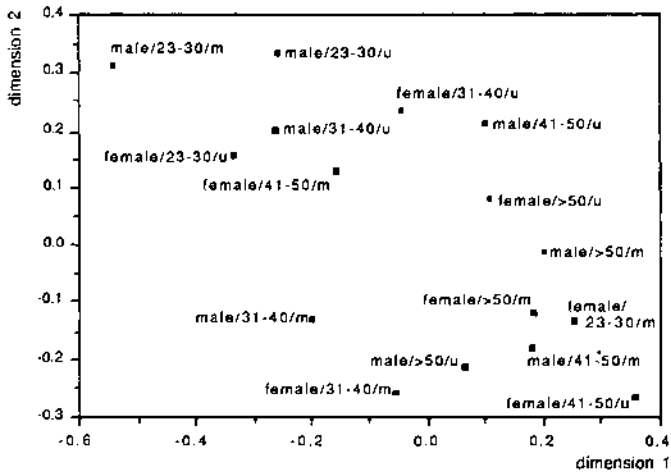

FIGURE 4 Supplementary points in two dimensions.

Adjacent time points are dependent, while times 1 and 3 are independent given 2, and hence the marginal matrix f_{i+k} can be derived as $f_{i+k} = \Sigma_j f_{ij+} f_{+jk} / f_{+j+}$. If the data in the three-way matrix can be adequately described by a *stationary* first-order Markov chain, then we need only one transition matrix and the initial proportions at time 1 to describe the data. For the three-wave example, this transition matrix can be estimated using the marginal tables with elements f_{ij+} and f_{+jk}.

The Burt matrix uses all possible transition matrices for two time points. MCA ignores three- and higher-way transition matrices. If the data follow a *nonstationary* or *stationary* first-order Markov chain, then all the information necessary to describe this chain is present in the Burt matrix, and MCA does not ignore any useful information. The sufficient margins are in the submatrices adjacent to the diagonal matrices. The other matrices in the Burt matrix can be derived from these sufficient margins. For the *stationary* Burt matrix, we find that the submatrix farthest from the diagonal (in Table 3, this is the matrix with elements f_{i+k}) approximates independence the most. This follows from the limiting-state behavior of Markov chains: If the time passed is taken to be large enough, the initial state of an object provides no information about the present state. A transition matrix from time 1 to the last time point would show a matrix with equal rows, i.e., an independent matrix.

So the farther apart two time points are, the more unrelated they are. The implication for MCA is that in a solution, the intermediate time points, which are the time points that are most related to the largest number of other time points, will have the largest contribution to the solution. For *nonstationary* Markov chains, similar results can be shown to hold under some conditions.

5.4. *Multivariate Panel Data*

So far, we have discussed the analysis of univariate panel data. Multivariate panel data could be dealt with in this framework by *stacking* the categories of these variables into a new composite variable and by dealing with this variable as if it were an ordinary variable. Another possibility, on which we will focus in this section, is to code the multivariate panel data into the *three-way superindicator matrix* and to analyze this matrix by first recoding it into two-way matrices. This differs from the situation for univariate panel data in that we are now able to work with the *LONG* matrix, because we have more than one 1 in each row. So we now have *four* possible matrices that could be analyzed: (a) the *BROAD* matrix $\mathbf{G}^{(ij)}$, in which the T slices of the three-way data block are concatenated horizontally; (b) the *LONG* matrix $\mathbf{G}^{(ij)I}$, in which the T slices of the three-way data block are concatenated vertically; (c) the marginal matrix $\mathbf{G}^{II}$, which can be considered again as a constrained analysis of the *BROAD* matrix but which also gives a constrained analysis of the *LONG* matrix, in which the T sets of object scores are constrained to be identical; and (d) the marginal matrix $\mathbf{G}^{TJ}$, which consists of the column margins of the *BROAD* matrix.

The analysis of the *BROAD matrix* $\mathbf{G}^{i(jt)}$ can perhaps be most easily understood when we think of it as a PCA for nominal variables. For *quantitative* longitudinal data, it is not unusual to construct *BROAD* matrices (cf. Visser 1985) and to analyze them with PCA or factor analysis. Many types of information are to be found in the solution of the MCA of $\mathbf{G}^{i(jt)}$. First, it can show the relation between distinct variables at identical time points. Second, it can show the relation between identical variables at distinct time points. And third, it can show the relation between distinct variables at distinct time points. One type of information is not displayed in

the solution, namely, the relation between categories and time points. This information is in the *marginal matrix* $\mathbf{G}^{TJ}$ and has to be analyzed separately, as we analyzed the univariate data. The analysis of the *marginal matrix* $\mathbf{G}^{IJ}$ can again be considered as providing a constrained solution for the *BROAD* matrix—constrained in the sense that each category receives identical quantifications for all time points.

MCA of the *LONG matrix* gives only one set of scores for all the categories but T sets of scores for the objects, namely, one set for each time point. This analysis can also best be understood as a PCA for nominal variables. A possible drawback of MCA of the *LONG* matrix is that category quantifications can be found that distinguish in the first place the different time points and not the different individuals; i.e., there are T sets of object scores that are approximately similar in each set but very different between sets. This might happen when the distributions of the categories differ considerably over time points. Such a solution might be considered rather uninteresting, because a similar—and more easily interpretable—solution might be obtained from the analysis of the marginal matrix $\mathbf{G}^{TJ}$. (This gives a solution in which object scores at identical time points are restricted to be the same.)

In the context of quantitative variables, a similar problem is that average differences in the analysis of *LONG* matrices can lead to spurious correlations. This is sometimes circumvented by vertically concatenating T slices that are each in deviation from the mean (see Bentler 1973). Escofier (1988) recently proposed a comparable procedure, which she called conditional MCA (see also van der Heijden 1987, ch. 7). Using the generalizations of CA (see equation (7)), we decompose not the departure from independence for the total matrix $\mathbf{G}^{(it)j}$ but the departure from independence in each of the T submatrices. So, for $\mathbf{P}$ in equation (7), we take the observed *LONG* matrix $\mathbf{G}^{(it)j}$, whereas the matrix $\mathbf{Q}$ has elements $g_{(it)+}g_{(+t)j}/g_{(+t)+}$. When for $\mathbf{S}_r$ and $\mathbf{S}_c$ simply the margins of $\mathbf{P}$ are taken, it can be shown that the generalized CA solution can be obtained with ordinary CA programs by analyzing the matrix $(\mathbf{P}-\mathbf{Q}+\mathbf{E})$. In fact, conditional MCA can be used in a similar way in the more general situation in which we want subgroups of objects to have an average of zero (see also van der Heijden 1987 and van der Heijden and Meijerink 1989 for more details).

6. MULTIPLE CORRESPONDENCE ANALYSIS OF EVENT HISTORY DATA

Event history data specify not only *the sequence* of states (categories) that apply to an object but also *how long* each object is in a specific state. Event history data can be coded into the familiar three-way superindicator matrix **G** by going from continuous time to *discrete* time points and making a superindicator matrix for each time point. So, $g^m_{ijt}=1$ if object i falls into category j of variable m at time point t, and $g^m_{ijt}=0$ otherwise. In this approach, event history data are just a special form of panel data, namely, a form that tells us for each time point the category in which an object falls.

In general, the number of time points T will be very large; for example, $T=1{,}440$ if we consider one day divided into minutes. Therefore, it is difficult to analyze the data without imposing any restrictions. We restrict the analysis by aggregating the data over time points within larger periods. For example, we can subdivide a day into hours, and instead of 1,440 time points, we then construct a matrix $\tilde{\mathbf{G}}$ with 24 time periods p ($p=1, \ldots, P$), where elements $\tilde{g}^m_{ijp}$ denote the number of minutes in hour p that object i spent in category j of variable m. This idea was first worked out by Deville and Saporta (1980, 1983); see also Saporta (1981, 1985), de Leeuw et al. (1985), and van der Heijden (1987). Let's again consider the analysis of the *BROAD* matrix $\tilde{\mathbf{G}}^{I(JP)}$, the *LONG* matrix $\tilde{\mathbf{G}}^{(IP)J}$, and the two marginal matrices $\tilde{\mathbf{G}}^{IJ}$ and $\tilde{\mathbf{G}}^{PJ}$.

In principle, the analysis of these matrices does not differ much from the analysis of panel data. Therefore, we will not distinguish between the univariate and multivariate cases. The *BROAD* matrix $\tilde{\mathbf{G}}^{I(JP)}$ shows how the objects spend their time in the periods chosen. MCA of the *BROAD* matrix will show how objects depart from the average time spent. Unlike the analysis of $\mathbf{G}^{I(JT)}$, in the analysis of $\tilde{\mathbf{G}}^{I(JP)}$ the category quantifications are restricted to be identical within each period. It will be clear that to save information, we must choose the periods in such a way that *within* periods, objects do not change states much but that *between* periods, the distributions are as different as possible. This consideration can lead to a choice of periods of unequal length (see, e.g., de Leeuw et al. 1985).

The average from which the objects depart is collected in the column margins of the *BROAD* matrix, i.e., the matrix $\tilde{\mathbf{G}}^{PJ}$. This matrix shows that some states are better filled in some periods and that other states are better filled in others. An analysis of this matrix supplements the analysis of the *BROAD* matrix. In the marginal matrix $\tilde{\mathbf{G}}^{IJ}$, the number of periods is reduced to one. This matrix shows how the objects spend their time. In an MCA of $\tilde{\mathbf{G}}^{IJ}$, the category quantifications are restricted to be the same over the complete time range.

The situation for the *LONG* matrix is slightly different from what it was for univariate panel data: If we have event history data for one variable and if the periods are chosen so that state changes occur, the analysis of the *LONG* matrix will not give a trivial solution. In general, however, the analysis of the *LONG* matrix will approach the (uninteresting) analysis of a diagonal matrix if the number of periods is chosen to be sufficiently large, so that time periods will be short and the number of objects in only one state will be large.

6.1. *Example*

In the first section, we discussed the analysis of a matrix of 18 hours by 25 categories, with each cell representing the number of minutes in an hour that 326 couples spend in a specific category. This matrix can be considered as the margin $\tilde{\mathbf{G}}^{TJ}$ of the three-way data block $\tilde{\mathbf{G}}$ of order $326\times25\times18$. We will now analyze the *BROAD* matrix $\tilde{\mathbf{G}}^{I(JT)}$, having order $326\times(25\times18)$. In fact, the number of the columns of the matrix is slightly smaller than $25\times18=425$; it is 343. Not all joint activities are displayed at each hour (see zero frequencies in Table 1), and we find no quantifications for such joint activities.

MCA of the *BROAD* matrix gives as first few eigenvalues 0.422, 0.390, 0.339, 0.294, 0.271, 0.259, 0.249. A plot of the 326 objects (the couples) can be found in Figure 5. The cloud of points has a peculiar form. On the first dimension, a distinction is made between two groups: A large number of couples appears on the left, and a smaller number appears on the right. Low on the second dimension, an even smaller group of couples is visible. This phenomenon goes on in higher dimensions. In each dimension, a

TABLE 5
Category Quantifications on Dimension 1

	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
HH	0.04	0.31	1.34	1.89	2.07	2.11	1.64	1.80	1.97	1.88	1.19	0.28	0.07	0.00	−0.03	−0.03	−0.04	−0.04
HW	−0.31	−0.28	0.02	0.25	0.39	0.38	0.07	0.05	0.06	0.11	−0.09	−0.49	−0.45	−0.26	−0.11	−0.01	0.13	0.16
HS	—	—	—	—	2.19	2.23	2.13	—	2.07	—	2.06	0.06	−0.56	−0.27	—	—	—	—
HT	−0.20	−0.28	−0.06	0.94	1.85	1.38	0.82	0.67	0.68	0.70	−0.25	−0.32	−0.32	−0.23	−0.24	−0.45	0.16	0.38
HO	0.18	0.97	1.09	1.88	1.88	—	0.39	1.63	0.78	0.73	0.96	0.87	0.45	0.12	0.08	0.12	0.38	0.59
WH	−0.24	−0.46	−0.27	−0.09	−0.02	0.01	−0.20	−0.19	−0.19	−0.17	−0.24	−0.52	−0.44	−0.29	−0.45	−0.71	−0.69	−0.38
WW	−0.03	−0.61	−0.61	−0.62	−0.62	−0.61	−0.67	−0.68	−0.67	−0.68	−0.67	−0.66	−0.66	−0.69	−0.33	−0.30	−1.06	−1.17
WS	—	—	—	—	−0.56	−0.60	−0.28	−0.09	−0.16	−0.60	−0.78	−0.97	—	—	—	—	—	—
WT	−0.81	−0.63	−0.70	−0.52	−0.34	−0.26	−0.41	−0.42	0.05	−0.25	−0.62	−0.75	−0.90	0.30	−0.40	0.15	−0.64	—
WO	—	—	0.12	0.18	0.03	0.04	−0.37	−0.16	0.38	0.13	−0.40	−0.37	—	—	0.29	0.37	0.37	0.41
SH	—	—	—	2.13	2.02	2.04	1.03	1.26	1.58	0.99	1.09	0.32	0.64	−0.31	−0.21	−0.08	−0.08	—
SW	—	—	—	−0.18	0.00	−0.29	−0.60	−0.45	−0.28	−0.30	−0.39	−0.27	0.88	—	—	—	—	—
SS	—	—	—	1.99	2.06	1.52	0.38	−0.11	0.64	1.97	1.43	0.92	2.40	1.95	2.14	2.79	—	—
ST	—	—	—	—	1.78	1.78	−0.54	−0.63	−0.18	1.06	0.13	−0.52	—	—	—	—	—	—
SO	—	—	—	—	—	—	—	−0.71	—	1.95	—	−0.62	—	−0.61	—	—	—	—
TH	−0.52	−0.38	0.00	1.04	1.82	1.51	0.45	0.79	1.72	1.22	0.04	−0.21	−0.32	0.50	0.14	0.40	−0.28	−0.64
TW	—	−0.62	−0.58	−0.17	0.04	−0.19	−0.48	−0.41	−0.36	−0.38	−0.56	−0.52	−0.49	−0.53	−0.33	−0.06	0.40	—
TS	—	—	—	—	—	—	−0.44	2.13	−0.41	1.05	−0.93	−0.58	—	−1.03	—	—	—	—
TT	−0.58	−0.50	−0.32	0.54	1.39	1.68	0.39	0.58	1.30	1.01	−0.22	−0.24	0.03	0.18	0.09	0.52	0.95	−0.23
TO	—	—	—	—	1.88	—	−0.27	−0.64	1.92	−0.40	0.03	−0.29	−0.59	0.26	−0.37	−0.38	0.31	0.17
OH	—	2.10	2.40	1.26	−0.13	—	−0.32	0.55	2.13	1.47	0.91	1.15	0.94	0.65	0.48	0.36	0.59	0.76
OW	—	−0.93	−0.68	−0.40	0.11	0.37	−0.06	0.03	0.37	0.30	0.37	−0.04	−0.54	−0.50	−0.72	−0.79	−0.83	−0.70
OS	—	—	—	—	—	—	−0.65	—	2.13	—	0.41	−0.09	−0.09	−0.94	−1.03	—	—	—
OT	—	—	—	0.93	0.94	0.83	0.25	—	2.13	1.07	0.39	0.64	0.78	0.63	−0.79	−0.90	−0.61	−0.88
OO	—	—	—	—	1.78	1.76	1.60	1.46	1.12	1.39	1.69	1.20	0.52	0.47	0.34	0.34	0.24	0.52

Note. See note to Table 1.

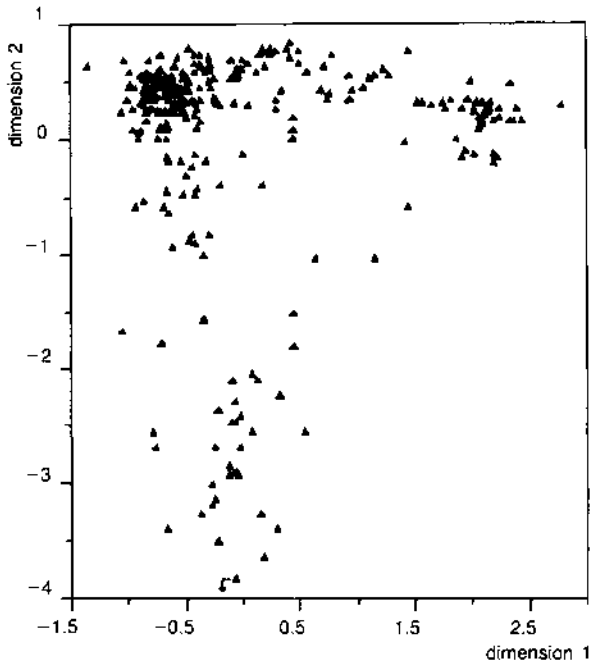

FIGURE 5 Row points in two dimensions.

small group of couples is separated from the rest: Couples in a separated cluster differ in the same respect from the average event history that is summarized in $\tilde{\mathbf{G}}^{TJ}$ and that is dominated by the larger group of couples. This is also reflected in the eigenvalues, which drop from 0.422 only very gradually. We conclude that the bulk of couples have quite the same event histories, probably because we selected couples in which both partners work more than 25 hours a week and we are only studying working days.

The category quantifications in Table 5 and their plot in Figure 6 show how the group on the right in Figure 5 differs from the larger group. In Figure 6, the hours are presented along the horizontal, and the quantifications on dimension 1 are presented along the vertical. For the five categories in which both partners perform the same activity, we have connected the points for adjacent hours. To determine the points on which we must concentrate, we study the contribution of each individual point to dimension 1;

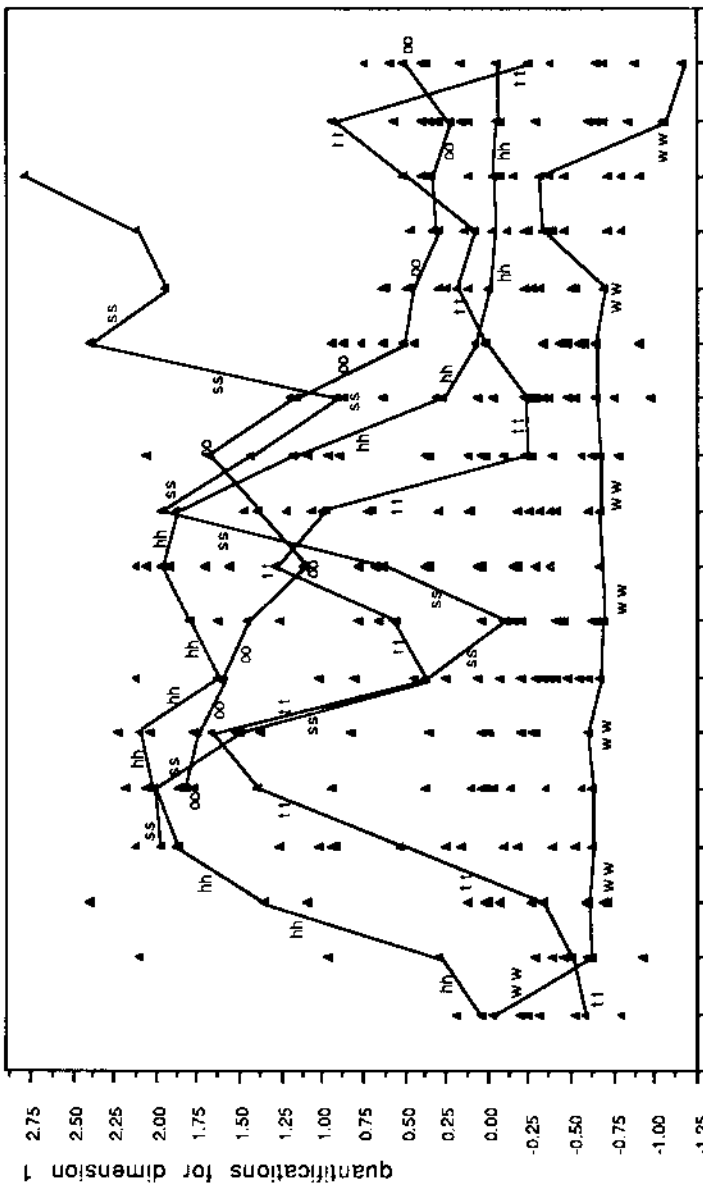

FIGURE 6 Category points. Quantifications on first dimension are set out vertically, the hours horizontally. Category points are connected only for identical activities of husbands and wives (i.e., *HH, WW, SS, TT, OO*). See Table 5 for labels of the other points.

thus, we find a subdivision (in proportions) of the eigenvalue 0.422. By adding up over the 18 hours, we find contributions of the 25 joint activities to dimension 1; by adding up over the 25 joint activities, we find contributions of the 18 hours on dimension 1. When we focus on the *hours*, we find that the hours between 8:00 A.M. and 5:00 P.M. clearly contribute most; thus, on dimension 1, the couples are distinguished mainly on the basis of their departure from the average joint activities in these hours, say working hours. When we then focus on the *joint activities*, we find that the *HH* category (both partners at home) and the *WW* category (both partners at work) contribute most (57 percent and 24 percent, respectively). So the couples on the right in Figure 5 are separated from the couples on the left mainly because they depart in opposite ways on these joint activities. Figure 6 shows that during working hours, the couples on the right are much more often both at home and much less often both at work than the couples on the left. Since we know from the sample that both partners work at least 25 hours a week, the couples on the right probably have a day off. After working hours, the couples on the right do not differ much from the couples on the left. When we study Table 5 and Figure 6 more carefully, we find that most of the joint activities in which at least one of the partners is working (i.e., one *W*) are quantified negatively in all hours; so for the objects on the left, at least one of the partners is working. These couples travel more often in the morning (*TT* has negative quantifications), probably to their work, and also more between 4:00 P.M. and 6:00 P.M., probably from work. We can also see a lunch dip for both partners travelling or shopping: In the morning and afternoon, these activities are performed more by the couples having a day off, whereas during lunch time, these behaviors are evenly spread between the couples on the right and those on the left. To interpret Figure 6, we first studied the activities that contribute much. We did not necessarily give much attention to the most extremely quantified activities, such as *SS* in the evening, since these activities may have a very low frequency. (*SS* is performed only 150 minutes after 6:00 P.M. However, the couples doing this are all quantified extremely positive.)

For the second dimension, we find that the hours between 8:00 A.M. and 4:00 P.M. contribute between 5 percent and 10

percent each, the evening hours contributing about 3 percent each. The activities contributing much are *WH* (58 percent) and *HW* (9 percent), *WO* (4 percent) and *OW* (5 percent), *WW* (5 percent), *SH* (5 percent) and *TH* (4 percent). A table similar to Table 5 (not displayed here for reasons of space) shows that the couples quantified negatively in Figure 5 are couples in which the wives work during ordinary working hours and the husbands are at home (*WH*) or doing other things (*WO*). During working hours wives also travel more while the husbands are at home (*TH*). For the rest of these couples, we find that the husbands work more than average in the evening while the wives are at home (*HW*) or doing other things (*OW*) and that wives shop more in the early evening while the husbands are at home (*SH*). We can classify the couples quantified negatively on dimension 2 roughly as couples in which the wives work more than average during ordinary working hours while the husbands do things other than working and the husbands work more in the evening while the wives do things other than working.

As a last step, we could study the relation between supplementary information and the object scores in the same way we did for panel data (compare Figure 4). Because of lack of space, we did not do this (but see van der Heijden 1987).

In sum, we think that bivariate and multivariate event histories can very well be studied by means of MCA. A possible limitation is the number of objects compared with the number of activities. If the former is small compared with the latter, the solution can become unstable. In general, however, solutions can be made more stable by restricting the quantifications of each category to be identical within larger time periods. This amounts to separating the total time range into only a few time periods, thus reducing the number of columns of the *BROAD* matrix.

6.2. *MCA Compared with Statistical Event History Analysis*

We now compare MCA of event history analysis with statistical approaches, such as those summarized in Allison (1984) and Tuma and Hannan (1984). Allison discusses a number of dimensions along which the statistical approaches differ, and we will use some of these to place MCA into his framework. A first

dimension is that between *regression methods* and *distributional methods*. In MCA we use both aspects in an analysis. First, homogeneous groups of objects are formed on the basis of the differences in distributions of behavior over time. This gives us one set of object scores for each dimension. Second, supplementary information is related to these sets of object scores by computing averages over supplementary categories. In MCA, we use a two-step procedure to find object scores and then relate these to the supplementary information. This is done in one step in the methods described by Allison. It is also possible to do this in a combined step for MCA using canonical CA (ter Braak 1986). With canonical CA we can find object scores that are restricted to be a linear combination of a set of supplementary variables. However, this has not yet been applied to the analysis of event history data.

A second dimension discussed in Allison is that of *repeated* versus *nonrepeated* events, and a third dimension is that of *single* versus *multiple kinds* of events. *Single nonrepeated events* can be dealt with in MCA by using in our three-way indicator matrix two categories: "event did not yet happen" and "event happened." Similarly, "censored" can be defined as a separate category as an alternative to "event happened." (Another approach using simple CA is proposed by Nakache, Asselain, and Lasry [1984] and will be discussed below.) In our approach, *single repeated events* can be dealt with by counting the number of times an event maximally happened, x, and defining $x+1$ categories, namely, "event did not yet happen," "event has happened once," to "event has happened x times." In fact, this is a way to analyze point processes with MCA. We deal with point processes here not by defining the points themselves but by defining the time between the points as distinct states. If the number of times an event happens for an object is very large and the number of objects to which this happens is small, MCA can form homogeneous groups of objects by placing this small number of objects in the periphery of the plot of object scores. However, this can be remedied by defining the state "event happened more than x times" in such a way that the number of objects in it is large enough to reduce the number of categories. When there are *multiple kinds of events*, we can deal with those in the same way, namely, by defining states an object is in. So, if we deal with unrepeated multiple kinds of events and there are, say,

K kinds of events, we define $K+1$ categories, namely, "nothing has happened yet" to "state K has happened." With *repeated multiple kinds of events* where a specific kind of event can follow itself, we define categories like "nothing has happened," "event k has happened one time," "event k has happened two times," "event k^* has happened one time," etc. When we deal with repeated multiple kinds of events where a specific event *cannot* follow itself, we can count the number of times events have happened, but we can also define only K possible categories, indicating the state an object is in. In fact, this is the case for the example that we analyzed in this paper, and it is also the only type of example that is published (see Deville and Saporta 1983; de Leeuw et al. 1985; van der Heijden 1987). However, we have practical experience with some of the other cases mentioned above, where MCA gave satisfactory results.

A fourth dimension mentioned by Allison (1984) is that of *discrete* time versus *continuous* time. MCA is applicable in both cases: In the case of discrete time, we work with panel data, and in the case of continuous time, we use the fact that continuous time is always measured with discrete precision, defining the state an object is in at a specific time point. Because the time dimension of the three-way (super)indicator matrix becomes very large when precise measurements are used, we reduce this matrix by defining time periods.

Another problem occurs when objects are not measured over the *same time range* (apart from the fact that, of course, the time scale can be defined in various ways). This situation can be dealt with by defining an extra category, "not observed yet" or "not observed anymore." However, the drawback is that MCA can find a solution that distinguishes objects on the basis of these categories. Therefore, it is probably better to define a person as missing during the period he/she is not observed and to code all categories in this period as 0. We have not used this to analyze event history data, but this approach to missing data works quite well in the context of MCA of "ordinary" incomplete data.

Nakache et al. (1984) perform simple CA on a special contingency table to analyze *unrepeated single event histories*. This work is similar to the work of Laird and Olivier (1981), who also use contingency tables to show that survival analysis can be performed using loglinear models. Nakache et al. construct a matrix in which the columns are the categories of the explanatory variables

and the rows are specific grouped lengths of survival times, split up for censored individuals and noncensored individuals. In the example they use, there are ten rows: five for censored persons and five for noncensored persons. The five survival lengths are 0–6 months, 6–24 months, 24–60 months, 60–120 months, and more than 120 months. In each cell of the matrix, we find the number of individuals who fall into a specific category of an explanatory variable and who have a specific survival time. This matrix is analyzed with simple CA. For the example Nakache et al. use, the survival time periods are quantified monotonically on the first dimension, both for the censored persons and for the noncensored persons. The profiles of the explanatory variables for the noncensored individual objects are fitted into this solution (this can be done with transition formula (3), where $\mathbf{C}$ is the matrix with category scores for the explanatory variables and $\mathbf{D}_r^{-1}\mathbf{P}$ is the matrix with profiles for the individuals), thus obtaining scores for each individual. Nakache et al. correlated these scores with the scores for these noncensored individuals using the Cox regression model and found a correlation of 0.88, which gives some justification for the exploratory procedure they applied.

As we have tried to describe above, MCA of event history data is a very flexible exploratory approach. Much research in this area is needed. For example, practical experience of MCA with many types of event history data is limited, and it is not clear how or whether exploratory CA can be used prior to confirmatory survival analysis. In principle, it seems that it can be used for all sorts of event histories, and it can easily handle data with a large set of states (in our example, we used 25). Computationally, neither a large amount of objects nor a large amount of time periods creates a problem. However, results might become unstable when the number of time periods and categories is too small compared with the number of objects. This problem can be solved by making the periods larger, thus making the number of columns smaller, thus restricting the quantifications of categories to be constant for a longer time span.

7. CONCLUSIONS

We have discussed some of the properties of CA and MCA, concentrating on MCA of panel and event history data. Compared

with more usual data analysis approaches, MCA is less hampered by empty-cell problems because it uses only the bivariate marginal dependence in the data. Therefore, it can analyze data sets with many variables, categories, and time points that cannot be analyzed with approaches like Markov chain models or processes. On the other hand, MCA sometimes ignores information that is too important to ignore. However, we have seen that if a first-order Markov model fits adequately, MCA uses all information in the analysis.

Another difference between MCA and more usual approaches is that MCA finds scores for objects; therefore, it is possible to study *individual differences*. This is rather different from the more usual approaches, in which objects are treated as replications and differences between objects are usually neglected or considered to be the result of random error. There, if the differences between objects become too large, objects are often split up into a few subgroups and are considered again as replications in these subgroups. This is one of the ways to link supplementary information (like sex, age, SES) to the process under study. It is not possible to link all these supplementary variables at the same time because of empty-cell problems. On the other hand, in MCA we find quantifications for objects that emphasize differences between them. To these scores we can easily relate a large number of supplementary variables by computing averages for each category.

We think that MCA can be useful in obtaining a first idea of what is going on in the data. No assumptions have to be fulfilled for the technique to work. If we concentrate on the bivariate margins, interesting information might be ignored, but the range of possible applications becomes very large. Bivariate margins can be reasonably filled even though the total number of objects is much smaller than the total number of possible profiles (i.e., cells in the multiway contingency table). This is one important advantage of MCA. MCA is also a good alternative to many modeling techniques when the number of categories of the variables is large, making interpretation of parameters quite difficult. In this sense also, the MCA approach is far more exploratory in nature and we think in many applications also far more realistic, especially when no theory is available to use a specific model or distribution.

REFERENCES

Allison, P. D. 1984. *Event History Analysis*. Beverly Hills: Sage.

Bentler, P. M. 1973. "Assessment of Developmental Factor Change at the Individual and the Group Level." Pp. 145–74 in *Life-Span Developmental Psychology: Methodological Issues*, edited by J. R. Nesselroade and H. W. Reese. New York: Academic Press.

Benzecri, J. P. et collaborateurs. 1973. *Analyse des donnees*. 2 vols. Paris: Dunod.

Bishop, Y. M. M., S. E. Fienberg, and P. W. Holland. 1975. *Discrete Multivariate Analysis*. Cambridge, MA: MIT Press.

Bock, R. D. 1960. *Methods and Applications of Optimal Scaling*. Research Memorandum No. 25. Chapel Hill: University of North Carolina, Psychometric Laboratory.

de Leeuw, J. 1973. "Canonical Analysis of Categorical Data." Ph.D. diss., University of Leiden.

———. 1983. "On the Prehistory of Correspondence Analysis." *Statistica Neerlandica* 37: 161–64.

de Leeuw, J., and P. G. M. van der Heijden. 1988. "Correspondence Analysis of Incomplete Tables." *Psychometrika* 53: 223–33.

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.

Deville, J.-C., and G. Saporta. 1980. "Analyse harmonique qualitative." Pp. 375–89 in *Data Analysis and Informatics*, edited by E. Diday. Amsterdam: North Holland.

———. 1983. "Correspondence Analysis, with an Extension Towards Nominal Time Series." *Journal of Econometrics* 22: 169–89.

Escofier, B. 1984. "Analyse factorielle en reférénce à un modèle: Application à l'analyse de tableaux d'échanges." *Revue de Statistique Appliquée* 32: 25–36.

———. 1988. "Analyse des correspondances multiples conditionelles." Pp. 333–42 in *Data Analysis and Informatics* 5, edited by E. Diday. Amsterdam: North Holland.

Gifi, A. 1981. *Nonlinear Multivariate Analysis*. Leiden: University of Leiden, Department of Datatheory.

Gilula, Z., and S. J. Haberman. 1986. "Canonical Analysis of Contingency Tables by Maximum Likelihood." *Journal of the American Statistical Association* 81: 780–88.

Goodman, L. A. 1981. "Association Models and Canonical Correlation in the Analysis of Cross-Classifications Having Ordered Categories." *Journal of the American Statistical Association* 76: 320–44.

———. 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.

———————. 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*. New York: Academic Press.

Kendall, M. G., and A. Stuart. 1967. *The Advanced Theory of Statistics*. 2d ed. London: Griffin.

Laird, N., and D. Olivier. 1981. "Covariance Analysis of Censored Survival Data Using Log-Linear Analysis Techniques." *Journal of the American Statistical Association* 76: 231–40.

Lebart, L., A. Morineau, and K. M. Warwick. 1984. *Multivariate Descriptive Statistical Analysis: Correspondence Analysis and Related Techniques for Large Matrices*. New York: Wiley.

Meerens, H., L. Boer, and A. Tan. 1988. *Gevolgen inkomensachteruitgang*. Leiden: University of Leiden, Werkgroep A&W.

Meulman, J. 1982. *Homogeneity Analysis of Incomplete Data*. Leiden: D.S.W.O. Press.

Moning, H. 1983. *The National Travel Survey in the Netherlands*. Heerlen: Central Bureau of Statistics.

Nakache, J.-P., B. Asselain, and C. Lasry. 1984. "Contribution de l'analyse factorielle des correspondances à l'etude multidimensionelles de données tronqués." Pp. 99–108 in *Data Analysis and Informatics 3*, edited by E. Diday. Amsterdam: North Holland.

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

———————. 1986. "Classification with a Variety of Categorical Data." Pp. 353–59 in *Classification as a Tool of Research*, edited by W. Gaul and M. Schrader. Amsterdam: North Holland.

Plewis, I. 1985. *Analyzing Change: Measurement and Explanation Using Longitudinal Data*. New York: Wiley.

Saporta, G. 1981. "Methodes exploratoires d'analyse de donnees temporelles." *Cahiers du Buro* 37–38.

———————. 1985. "Data Analysis for Numerical and Categorical Individual Time-Series." *Applied Stochastic Models and Data Analysis* 1: 109–19.

Tenenhaus, M., and F. W. Young. 1985. "An Analysis and Synthesis of Multiple Correspondence Analysis, Optimal Scaling, Dual Scaling, Homogeneity Analysis, and Other Methods for Quantifying Categorical Multivariate Data." *Psychometrika* 50: 91–119.

ter Braak, C. J. F. 1986. "Canonical Correspondence Analysis." *Ecology* 67: 1167–79.

Tuma, N. B., and M. T. Hannan. 1984. *Social Dynamics: Models and Methods*. New York: Academic Press.

van der Heijden, P. G. M. 1987. *Correspondence Analysis of Longitudinal Categorical Data*. Leiden: D.S.W.O. Press.

van der Heijden, P. G. M., and J. de Leeuw. 1985. "Correspondence Analysis Used Complementary to Loglinear Analysis." *Psychometrika* 50: 429–47.

van der Heijden, P. G. M., A. de Falguerolles, and J. de Leeuw. 1989. "A Combined Approach to Contingency Table Analysis Using Correspondence Analysis and Loglinear Analysis." *Applied Statistics*, forthcoming.

van der Heijden, P. G. M., and F. Meijerink. 1989. "Generalized Correspondence Analysis of Multiway Contingency Tables and Multiway (Super-) Indicator Matrices." In *The Analysis of Multiway Data Matrices*, edited by R. Coppi and S. Bolasco. Amsterdam: North-Holland.

Visser, R. A. 1985. *The Analysis of Longitudinal Data in Behavioural and Social Research*. Leiden: D.S.W.O. Press.