Computational Statistics and Data Analysis 53 (2009) 3209–3216

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Trend vector models for the analysis of change in continuous time for multiple groups

Mark de Rooij *

Leiden University Institute for Psychological Research, Methodology and Statistics Unit, Leiden, The Netherlands

ARTICLE INFO

Article history:
Available online 9 October 2008

ABSTRACT

A problem with the modeling of repeated multinomial response data is the dimensionality of the response variable. For reducing this dimensionality and enhancing interpretability multidimensional scaling techniques are utilized. The resulting trend vector model provides an easily interpretable graphical display with trajectories of different groups over time. A generalized estimating equations scheme is employed for obtaining estimates of the parameters. Model selection is based on the Bayesian Information Criterion and the bootstrap. For illustration, the model is applied to a data set.

1. Introduction

Three main families of models exist for the analysis of longitudinal data (Diggle et al., 2002; Molenberghs and Verbeke, 2005): marginal models, transitional models, and subject specific models. In the first type, marginal models, responses are modeled marginalized over all other responses; the association structure is typically captured by a set of association parameters. In transitional models any response in the sequence is modeled conditional upon (a subset of) past responses. In subject specific models the responses are assumed independent given a set of subject specific parameters. The three types of models typically answer different questions. The marginal approach handles the question how, on average, the system of probabilities evolves over time in the population, whereas the subject specific approach handles the question how this system of probabilities evolves over time for each individual subject. The transitional approach takes into account the response at previous time points which is qualitatively different from the two other perspectives. Whereas in the case of a normally distributed outcome variable these types of models are naturally connected, for categorical outcomes there is no close connection (Diggle et al., 2002; Molenberghs and Verbeke, 2005). For most kinds of data well-established statistical tools have been developed for the analysis of longitudinal data in each of the three families. For overviews, see for example Diggle et al. (2002), Verbeke and Molenberghs (2000), Molenberghs and Verbeke (2005) or Hedeker and Gibbons (2006). A type of data that has received less attention is repeated multinomial unordered data, although there have been some developments (Hartzel et al., 2001; Hedeker and Gibbons, 2006; Lipsitz et al., 1994). The major problem with this kind of data is the dimensionality, i.e. when the response variable has G classes the dimensionality is $G - 1$ in multinomial logistic regression models, whereas in most statistical models for continuous or binary response variables the dimensionality is 1. This increased dimensionality may trouble estimation and interpretation.

Multidimensional scaling procedures have been proposed for the analysis of repeated multinomial data that reduce the dimensionality. De Rooij (2008a), for example, showed distance models for two-way and three-way transition frequency data with a 'law of gravity' interpretation. Simpler representations of trends are given in the so-called slide vector model (Zielman and Heiser, 1993; De Rooij and Heiser, 2000), which simultaneously provide a similarity structure of the

* Corresponding address: Leiden University Institute for Psychological Research, Methodology and Statistics group. P.O. Box 9555, 2300 RB Leiden, The Netherlands. Tel.: +31 71 5274102.
E-mail address: rooijm@fsw.leidenuniv.nl.

0167-9473/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2008.09.030

Table 1
The data format.

Subject	Time	Response				Explanatory variables			
1	1	f_{111}	f_{112}	$\dots$	f_{11G}	x_{111}	x_{112}	x_{113}	$\dots$
1	2	f_{121}	f_{122}	$\dots$	f_{12G}	x_{121}	x_{122}	x_{123}	$\dots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
i	t	f_{it1}	f_{it2}	$\dots$	f_{itG}	x_{it1}	x_{it2}	x_{it3}	$\dots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

outcome classes as well as a trend. The outcome of a slide vector model is easily interpretable as demonstrated in the two papers referred to above. However, both the models developed in De Rooij (2008a) as well as the slide vector models are restricted to discrete time problems with two or three measurement occasions. Having more measurement occasions or when measurements are not taken at the same time points for different subjects makes application of these methods impossible.

One major problem with longitudinal data is that subjects drop out of the study. Both, in the law of gravity models of De Rooij (2008a) as well as in the slide vector models transition frequency data are analyzed. Having such data it is difficult to represent cases with incomplete data. One way to deal with this problem is to adapt the transition frequency table by including an extra category labeled 'missing'. A problem with this approach is that the subjects with missing data at a specific time point are considered to be a homogeneous group (Meulman, 1982). A solution to this problem is to give each missing subject a separate category, but this would make the frequency table very sparse and parameter estimates ill defined.

In many applied longitudinal problems there are different groups and the main goal of the analysis is the difference in development of the groups. Although Zielman and Heiser (1993) point out a way to deal with multiple groups at two time points, their approach is limited since the trend over time has the same direction for all groups, and can only be stronger for certain groups compared to others. Moreover, when there are many groups their procedure may become problematic, since for each group a transition frequency table has to be defined, which will become very sparse for increasing number of groups.

In this paper a statistical model based on multidimensional scaling techniques for repeated multinomial data will be developed that can handle

(1) more than three time points;
(2) continuous time measurements;
(3) differing number of repeated measurements for the subjects;
(4) different trends over time for subgroups.

In Section 2 we will introduce our model, discuss visualization, estimation, and model selection. The proposed model will be applied to an empirical data set in Section 3 and we will conclude with some discussion.

2. Modeling of longitudinal multinomial data

2.1. The data

Before we introduce our model we show the data format and introduce notation. For $i = 1, \dots, n$ subjects we have for each time point $t = 1, \dots, T_i$ a response on a categorical variable with G categories, indexed by $g = 1, \dots, G$. This response is gathered in a dummy coded vector $\mathbf{f}_{it} = [f_{it1}, \dots, f_{itG}]^\top$ with $\sum_g f_{itg} = 1$. We also have for each subject at every time point a vector with explanatory variables, denoted by $\mathbf{x}_{it} = (x_{it1}, \dots, x_{itp})^\top$. In Table 1 we show the layout of the data. If a subject has a missing response for a specific time point it just means that the row corresponding to that subject and time point is deleted. Note that in Table 1 we have a single column with time information. Later we will include time into the vector of explanatory variables and transform it to dummy coding or maybe use time and squared time as explanatory variables.

2.2. The trend vector model

We will model the conditional probability $\pi_{gt}(\mathbf{x}_{it})$ of an outcome category $g = 1, \dots, G$ at time point $t = 1, \dots, T_i$, for a subject i ($i = 1, \dots, n$) with p-dimensional covariate vector $\mathbf{x}_{it}$. This vector contains both time and group information and possibly other explanatory variables. Note that the time variable in Table 1 is now included in this vector. The conditional probability will be modeled by the squared distance between two points in Euclidean space of dimensionality M ($M \leq G - 1$). Therefore we introduce ideal points $\mathbf{y}_{it}$ for the subjects and class-points $\mathbf{z}_g$ for the categories. The ideal points $\mathbf{y}_{it} = (y_{it1}, \ldots, y_{itM})^\mathsf{T}$, which are gathered in a matrix $\mathbf{Y} = (\mathbf{y}_{11}, \ldots, \mathbf{y}_{1T_1}, \ldots, \mathbf{y}_{nT_n})^\mathsf{T}$, are a linear combination of the predictor variables $\mathbf{X}$, i.e.,

$$\mathbf{Y} = \mathbf{X}\mathbf{B},$$

where $\mathbf{X} = (\mathbf{x}_{11}, \ldots, \mathbf{x}_{1T_1}, \ldots, \mathbf{x}_{nT_n})^\mathsf{T}$. The conditional probability that subject i at time point t will be in class g is then equal to

$$\pi_{gt}(\mathbf{x}_{it}) = \frac{\exp(-d^2_{(it)(g)})}{\sum_k \exp(-d^2_{(it)(k)})},$$

where $d^2_{(it)(g)}$ is the squared Euclidean distance between the ideal point for subject i at time point t and the class point for category g in M-dimensional space, i.e.

$$d^2_{(it)(g)} = \sum_{m=1}^{M} (y_{itm} - z_{gm})^2.$$

This model is coined the trend vector model. For interpretation the log odds of choosing category a above b at time point t for a given ideal point can be derived, namely

$$\log \frac{\pi_{at}(\mathbf{x}_{it})}{\pi_{bt}(\mathbf{x}_{it})} = d^2_{(it)(b)} - d^2_{(it)(a)}.$$

The odds are in favor of the category closest to the ideal point of a person: a clear and natural interpretation of the graphical display.

2.3. Estimation

The model we specified in Section 2.2 is a marginal model, it focuses on the marginal distributions at the different time points. For marginal models there are generally two estimation approaches to obtain parameters (see Agresti (2002), Chapter 11): maximum likelihood estimation and estimation using generalized estimating equations (GEE).

For maximum likelihood estimation the complete joint distribution has to be specified while the model only applies to the marginal distributions of the responses at the different time points. When the number of time points is large or when explanatory variables are continuous maximum likelihood estimation is computationally infeasible.

We will estimate our model by maximizing

$$L = \sum_{i=1}^{n} \sum_{t=1}^{T_i} \log \prod_{g=1}^{G} \pi_{gt}(\mathbf{x}_{it})^{f_{itg}}, \tag{1}$$

which is the likelihood function for cross-sectional data. In our case it is not a true likelihood, since the dependencies among the repeated responses are not taken into account. As is shown by Liang and Zeger (1986) maximizing L with repeated measurements does provide consistent estimates of the model parameters. However, standard errors computed from the Hessian or information matrix of this function are generally biased. To deal with this bias Liang and Zeger (1986) introduced a sandwich estimator. For generalized linear models Liang and Zeger (1986) also adapt the estimation equations using these sandwich function to obtain generalized estimating equations. Various forms of correlation structures have been proposed to obtain the sandwich function like independence, exchangeable, first order auto regressive, or unstructured. When maximizing (1) we implicitly use the GEE framework with independence assumptions to estimate the model parameters.

2.4. Identification

The parameters of the trend vector model are the regression weights $\mathbf{B}$ and the class points $\mathbf{Z}$. The model has rotational freedom and a more intricate indeterminacy, that is, the probabilities remain the same when a constant is added for each subject:

$$\pi_{gt}(\mathbf{x}_{it}) = \frac{\exp(-d^2_{(it)(g)})}{\sum_k \exp(-d^2_{(it)(k)})} = \frac{\exp(-d^2_{(it)(g)}) + c_i}{\sum_k \exp(-d^2_{(it)(k)}) + c_i}.$$

So, we can add a constant to each subjects' squared distances to the class points without changing the probabilities (see De Rooij (2008b)). Since the probabilities in our model are solely based on squared Euclidean distances, we have that a model based on any squared distance matrix $\mathbf{D}_*$ defined as $\mathbf{D}_* = \mathbf{D} + \mathbf{c}\mathbf{1}^\mathsf{T}$ provides the same probabilities as the model defined with squared distances $\mathbf{D}$. Suppose $\mathbf{B}$ and $\mathbf{Z}$ give $\mathbf{D}$ and $\mathbf{B}_*$ and $\mathbf{Z}_*$ give $\mathbf{D}_*$. How are these related? The squared distance matrices can be written as

$$\mathbf{D} = \text{diag}(\mathbf{X}\mathbf{B}\mathbf{B}^\mathsf{T}\mathbf{X}^\mathsf{T})\mathbf{1}^\mathsf{T} + \mathbf{1}(\text{diag}(\mathbf{Z}\mathbf{Z}^\mathsf{T}))^\mathsf{T} - 2\mathbf{X}\mathbf{B}\mathbf{Z}^\mathsf{T},$$

and

$$\mathbf{D}_* = \text{diag}(\mathbf{X}\mathbf{B}_*\mathbf{B}_*^\top\mathbf{X}^\top)\mathbf{1}^\top + \mathbf{1}(\text{diag}(\mathbf{Z}_*\mathbf{Z}_*^\top))^\top - 2\mathbf{X}\mathbf{B}_*\mathbf{Z}_*^\top.$$

Since these two are equal up to an additive row constant, it follows that

(1) $\text{diag}(\mathbf{X}\mathbf{B}\mathbf{B}^\top\mathbf{X}^\top)$ may change without restrictions;
(2) $\text{diag}(\mathbf{Z}_*(\mathbf{Z}_*)^\top) = \text{diag}(\mathbf{Z}\mathbf{Z}^\top) + q\mathbf{1}$ for any q.
(3) $\mathbf{X}\mathbf{B}\mathbf{Z}^\top = \mathbf{X}\mathbf{B}_*\mathbf{Z}_*^\top + \mathbf{c}\mathbf{1}^\top$.

Which means that we can transform $\mathbf{Z}$ and $\mathbf{B}$ to

$$\mathbf{Z}_* = \mathbf{1}\mathbf{v}^\top + \mathbf{Z}\mathbf{T}, \quad \text{and}$$
$$\mathbf{B}_* = \mathbf{B}(\mathbf{T}^{-1})^\top,$$

under the restriction that $\text{diag}(\mathbf{Z}_*(\mathbf{Z}_*)^\top) = \text{diag}(\mathbf{Z}\mathbf{Z}^\top) + q\mathbf{1}$. From these equations we see that a rotation is always possible, in that case $\mathbf{v} = \mathbf{0}$. Furthermore, in dimensionality $M = G - 1$, the number of indeterminacies is $M(M+1) - (G-1) = M^2$: any non-singular $\mathbf{T}$ can be used and this can be solved by finding an appropriate vector $\mathbf{v}$ such that the restrictions are true. Summarizing, we have M^2 unknowns in $\mathbf{T}$, M in $\mathbf{v}$ but with $G - 1$ restrictions. The number of indeterminacies thus equals $\max(M(M-1)/2, M(M+1) - (G-1))$, and the number of independent parameters is

$$\text{npar} = (p + G)M - \max(M(M-1)/2, M(M+1) - (G-1)). \tag{2}$$

In order to obtain an identified solution we observe that row-wise centering makes solutions equal. Moreover, if we define $\boldsymbol{\Pi} = \{\pi_{gt}(\mathbf{x}_{it})\}$, and $\boldsymbol{\Delta} = \log \boldsymbol{\Pi}$ we also have $-\boldsymbol{\Delta}\mathbf{J} = \mathbf{D}\mathbf{J}$, with $\mathbf{J} = \mathbf{I}_G - \mathbf{1}_G\mathbf{1}_G^\top/G$. This makes it possible to use the metric unfolding with single centering (Heiser, 1981; De Rooij, 2008b) for identification, since this solution ignores the row means altogether (Heiser, 1981, p. 53). This procedure works fine, except in the situation of maximum dimensionality, i.e. $M = G - 1$. In this case we identify the solution by a transformation of $\mathbf{Y}$ such that $\mathbf{Y}^\top\mathbf{Y} = n\mathbf{I}$ (which can be obtained using a singular value decomposition), and solve for $\mathbf{v}$.

2.5. *Model selection*

Model selection comprises two issues: determination of the dimensionality and determination of valuable predictors. For determining the dimensionality we use the Bayesian Information Criterion (BIC) for which Pan and Le (2001) showed that in a GEE scheme with working independence assumption (the scheme we use) it performed well for correlated data. The BIC is then defined as follows

$$\text{BIC} = -2L + \text{npar} \times \log(n),$$

with L defined in Eq. (1) and the number of parameters (npar) is defined in Eq. (2).

After choosing a specific dimensionality the solution will be bootstrapped to obtain empirical confidence intervals for the regression weights and to verify which predictor variables contribute to the trend vectors. It is important that resampling is based on the level of the subjects to reflect the correlation structure of the data (Pan and Le, 2001; Sherman and le Cessie, 1997). The generated bootstrap replicates automatically retain the same dependence structure as the original data.

2.6. *Visualization of trends*

The results of the classification model described above could be represented in a joint plot reminiscent of the biplot (Gower and Hand, 1996) where the variables are depicted by vectors with length and direction determined through the regression weights. For longitudinal data, however, we would like to represent the trajectories over time. More specifically, suppose we have two groups and both linear and quadratic time are important explanatory variables plus their interactions with the grouping variable. Just depicting each of these variables in a joint plot does not provide the information we would like to get out of a longitudinal analysis. What we would like is the trajectories for the two groups over time, i.e. the combined effect of linear and quadratic time for the two groups separately. To visualize such trajectories we need the estimated ideal points and join these together over the time variable. The ideal points are defined by the regression weight vectors, and so these have to be transformed into trend vectors. The procedure differs slightly for discrete and continuous time models, yielding piecewise linear trends and smooth trends, respectively. The procedures will be described using an example, but of course they are more general.

Consider an example with two groups measured at four time points and there is an interaction between group and time. In that case we have a dummy for group (x_g) and three dummy variables for time (x_{t1}, x_{t2}, x_{t3}), as well as three dummy variables for the interaction of group and time (x_{gt1}, x_{gt2}, x_{gt3}). The dummy variables for time are all zero at the baseline measurement, $x_{t1} = 1$ for the second measurement, zero otherwise. For the third time point $x_{t2} = 1$, zero otherwise, and for the fourth measurement $x_{t3} = 1$, zero otherwise. The interaction dummies are products of the time and group dummies. The ideal points for dimension m are given by the following equation

$$y_{itm} = x_g b_{gm} + x_{t1} b_{t1m} + x_{t2} b_{t2m} + x_{t3} b_{t3m} + x_{gt1} b_{gt1m} + x_{gt2} b_{gt2m} + x_{gt3} b_{gt3m}.$$

From this it follows that the trend vector for the group with $x_g = 0$ is obtained by joining the origin with the end-points of $\mathbf{b}_{t1}$, $\mathbf{b}_{t2}$, and $\mathbf{b}_{t3}$, while for the group with $x_g = 1$ the trend vector is obtained by joining the point $\mathbf{b}_g$ with the end-points of the vectors defined by $\mathbf{b}_g + \mathbf{b}_{t1} + \mathbf{b}_{gt1}$, $\mathbf{b}_g + \mathbf{b}_{t2} + \mathbf{b}_{gt2}$, and $\mathbf{b}_g + \mathbf{b}_{t3} + \mathbf{b}_{gt3}$. An example of such a graphical display is given in Fig. 1, discussion of which follows in Section 3.2.

For continuous time we obtain smooth trend vectors. Again consider the two group example with four time points where in this case linear and quadratic time are used as predictors in the model. Again suppose that there exists an interaction between group and time. First define t to be a time variable in the range of the measurements in the data. Now define two time variables $x_t = t$ and $x_{t^2} = t^2$. Like before we define x_{gt} and x_{gt^2} to be the product of the group dummy and the time variables. The ideal points in this case are defined by

$$y_{itm} = x_g b_{gm} + x_1 b_{tm} + x_{t^2} b_{t^2 m} + x_{gt} b_{gtm} + x_{gt^2} b_{gt^2 m}.$$

The trend vector for the group with $x_g = 0$ starts at the origin and having $t \times \mathbf{b}_t + t^2 \times \mathbf{b}_{t^2}$; for the group with $x_g = 1$ the trend vector starts at $\mathbf{b}_g$ and follows $\mathbf{b}_g + t \times (\mathbf{b}_t + \mathbf{b}_{gt}) + t^2 \times (\mathbf{b}_{t^2} + \mathbf{b}_{gt^2})$. An example of such a display can be found in Fig. 2, discussion of which follows in Section 3.3. The procedures for discrete and continuous time can both be easily adapted for any number of groups and any number of time points.

2.7. Confidence bands for the trend vectors

Often it is of interest to see the variability in trend vectors, i.e. to have confidence bands around the trend vectors. Simultaneously it is of interest to see the variability in the class points. Both can be obtained from the bootstrap. For the trend vector we can, at several predefined values of the explanatory variables $\mathbf{x}_0$, compute $\hat{\mathbf{y}}_0^b = \mathbf{x}_0^{\mathsf{T}} \hat{\mathbf{B}}^b$, where $\mathbf{B}^b$ is the bth bootstrap replicate. Assuming normality of the bootstrap we can compute the covariance matrix of $\mathbf{y}_0^b$, that is $(\widehat{\text{cov}}(\hat{\mathbf{Y}}_0))$. Confidence intervals are given by the interior of the ellipse defined by the equation

$$\left[\mathbf{c} - \hat{\mathbf{y}}_0\right]^{\mathsf{T}} \widehat{\text{cov}}(\hat{\mathbf{Y}}_0)^{-1} \left[\mathbf{c} - \hat{\mathbf{y}}_0\right] = \chi,$$

where χ is the threshold at a desired probability level of a chi-squared variable with M degrees of freedom (the length of $\mathbf{y}_0$). Similarly, a confidence ellipse around class point $\hat{\mathbf{z}}_g$ is given by the equation

$$\left[\mathbf{c} - \hat{\mathbf{z}}_g\right]^{\mathsf{T}} \widehat{\text{cov}}(\hat{\mathbf{Z}})^{-1} \left[\mathbf{c} - \hat{\mathbf{z}}_g\right] = \chi.$$

2.8. Software

We implemented the models, bootstrap and graphical procedures in MATLAB (Mathworks, 2006). The programs can be obtained from the author upon request.

3. Application

3.1. McKinney Homeless Research Project

This data set is discussed in chapter 10 and 11 in the book by Hedeker and Gibbons (2006). The data can be obtained from the website of Donald Hedeker, accompanying the book. The aim of the McKinney Homeless Research Project in San Diego was to evaluate the effectiveness of using an incentive as a means of providing independent housing to homeless people with severe mental illness. Housing certificates were provided from the Department of Housing and Urban Development to local authorities in San Diego. These housing certificates are designed to make it possible for low income individuals to choose and obtain independent housing in the community. A sample of 361 individuals took part in this longitudinal study and were randomly assigned to the experimental or control condition. Eligibility for the project was restricted to individuals diagnosed with a severe and persistent mental illness who were either homeless or at high risk of becoming homeless at the start of the study. Individuals' housing status was classified at baseline and at 6, 12 and 24 month follow up. The focus will be on examining the effect of the incentive on repeated housing outcomes across time. Table 2 gives some characteristics of the data.

In Table 2 we see that for the control group most subjects initially live on the street. The proportion living on the street goes down in the first year but then increases again. The proportion of subjects living in a community center raises the first year but then declines, while the proportion of subjects that lives independently raises over the time of study. For the experimental group the proportion of subjects living on the street decreases the first half year after which it increases and stabilizes. The proportion of subjects living in a community center decreases the first year but then increases again. The proportion of subjects living independently increases rapidly but then declines. All trends except for the experimental living on the street, indicate that a quadratic treatment of time might capture the pattern in the data well.

In the following two subsections we will analyze this data set, first treating time in a discrete manner, afterwards as continuous. Note that for this data set maximizing (1) simplifies since we can sum the data over subjects in the same group.

Table 2
Housing condition across time by group: proportions and sample size.

Group	Status	Time point			
		Baseline	6	12	24
Control	Street	.555	.186	.089	.124
	Community	.339	.578	.582	.455
	Independent	.106	.236	.329	.421
	N	180	161	146	145
Experimental	Street	.442	.093	.121	.120
	Community	.414	.280	.146	.228
	Independent	.144	.627	.732	.652
	N	181	161	157	158

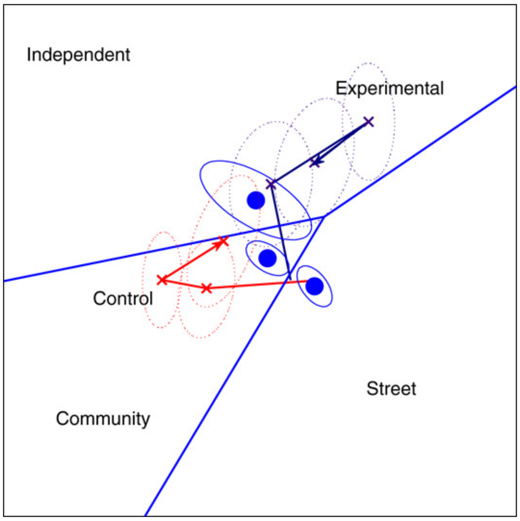

Fig. 1. Trend vector model for homeless data treating time discretely. The two arrows represent the trajectories for the experimental and control condition. The dotted ellipses give 1 standard deviation confidence intervals around the trend vectors at the points defined by the crosses. The solid ellipses around the class points give 1 standard deviation confidence regions. The lines represent boundaries where the odds for two categories are even.

3.2. Discrete treatment of the time variable

If we treat time discretely we have the following seven predictors variables

- (**G**) a dummy variable for experimental *vs.* control, $x_g = 1$ for experimental;
- (**T1**) dummy for 6 month *vs.* baseline, $x_{t1} = 1$ for $t = 6$, zero otherwise;
- (**T2**) dummy for 12 month *vs.* baseline, $x_{t2} = 1$ for $t = 12$, zero otherwise;
- (**T3**) dummy for 24 month *vs.* baseline, $x_{t3} = 1$ for $t = 24$, zero otherwise;
- (**GT1**) group by time (T1) interaction dummy, $x_{gt1} = x_g \times x_{t1}$;
- (**GT2**) group by time (T2) interaction dummy, $x_{gt2} = x_g \times x_{t2}$;
- (**GT3**) group by time (T3) interaction dummy, $x_{gt3} = x_g \times x_{t3}$.

Since the response variable has three categories the solution has either one or two dimensions. For the one-dimensional solution the BIC-statistic equals 2472.9 where it is 2447.3 for the two-dimensional solution. It seems that a two-dimensional solution is needed to represent this data. For the two-dimensional solution a balanced bootstrap was performed using 500 resamples. The results of the bootstrap (not shown) indicate that there were already initial differences between the control and experimental group and that both time and the interaction between time and condition resulted in regression weights different from zero.

The trend vector model with one standard deviation confidence bands is shown in Fig. 1. In this figure both the trajectories for the two groups as well as boundary lines are shown. These boundary lines show where the odds of choosing one of two categories are even. It can be seen that the experimental and control group have very different trajectories over time, i.e., the experimental group goes to independent housing much quicker than the control condition and remains in the independent housing class while the control group moves towards the community housing class and at the end makes a move in the direction of the independent housing class. At the end of the evaluation period the control group has about even odds of being in the independent housing class or the community housing class. The confidence bands around the trend vectors as well as the confidence regions of the class points indicate that the configuration found is quite stable.

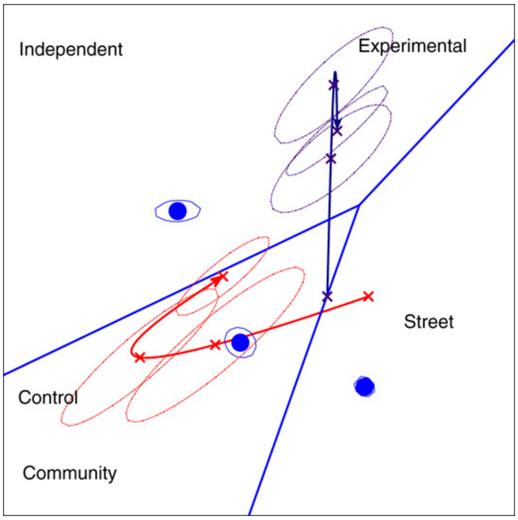

Fig. 2. Trend vector model for homeless data treating time continuously. The two arrows represent the trajectories for the experimental and control condition. The dotted ellipses give 1 standard deviation confidence intervals around the trend vectors at the points defined by the crosses. The ellipses around the class points give 1 standard deviation confidence regions. The lines represent boundaries where the odds for two categories are even.

3.3. Continuous treatment of the time variable

We will again model the homeless data but now treating time in a continuous manner. As discussed above in Table 2 it can be seen that most proportions go first up (or down) and later down (or up), which suggests a quadratic time pattern. Therefore the following predictors will be used

(**G**) a dummy variable for experimental *vs.* control, $x_g = 1$ for experimental;;
(**T**) linear time with values $x_t = t, t = 0, 6, 12, 24$;
($\mathbf{T}^2$) quadratic time with values $x_{t^2} = t^2, t = 0, 6, 12, 24$;
(**GT**) group by time interaction variable defined as $x_{gt} = x_g \times x_t$;
($\mathbf{GT}^2$) group by time interaction variable defined as $x_{gt^2} = x_g \times x_{t^2}$.

In one dimension the BIC equals 2470.9 while in two dimensions it is 2436.4. Again the two-dimensional solution is favored. The results of the bootstrap (not shown) indicate that all predictors are different from zero.

The resulting trend vector plot is shown in Fig. 2. It is very similar to the discrete time plot (Fig. 1). The experimental group is moving much faster into the independent housing class, while the control group is moving into the community housing group. The quadratic time effect for the experimental condition merely effects in a return in the direction of the starting position at the end of the study, while for the control condition it bends towards the independent housing class.

4. Discussion

A marginal model for longitudinal multinomial data was proposed that utilized multidimensional scaling techniques in order to reduce the dimensionality. Compared to other multidimensional scaling models our model is able to handle data where (1) people have different number of measurements; (2) measurements are obtained continuously over time; (3) subjects are classified in groups with possible different trends. This can all be done in the so-called trend vector model.

The trend vector model is an easily interpretable model for longitudinal multinomial data. It results in a graphical display that pictures the trend over time of the groups under study. In maximum dimensionality the trend vector model equals the multinomial logit model, and thus the trend vector model provides a graphical display of the multinomial logit model, in which trajectories are represented. These are of main interest in longitudinal research.

When treating time in a continuous manner with linear and quadratic time effects, the composite effect of time might be hard to interpret in a multinomial logit model, whereas in our graphical display the interplay between these two effects is easily interpreted. The trend vector model has the possibility of dimension reduction, which might be a virtue in case the number of classes in the response variable is large.

We estimate the model using the cross-sectional likelihood, ignoring the dependencies among the responses. This can be justified by GEE theory: we basically use GEE with working independence assumptions to obtain parameter estimates. Independence assumptions might seem a bit naive, but a fair amount of studies indicate that the more sophisticated assumptions might be counterproductive (see Crowder (1995), Lumley (1996), O'Hara Hines (1997), Sutradhar and Das (1999) and Pepe and Anderson (1994)). Confidence intervals for the parameter estimates were obtained using the bootstrap.

Alternatively we could have worked with the robust form as is often discussed in relation to GEE estimation. Further research is needed on standard errors and other issues in model selection.

We displayed one standard deviation confidence bands around the trend vectors. One should however, be cautious in interpreting such confidence intervals, since in multidimensional scaling techniques points often move together (see Kiers and Groenen (2006)). In our case this means that a trend vector moves but simultaneously the class point also moves. Visualizing such dependencies in a display is impossible, therefore Kiers and Groenen (2006) used movies of bootstrap solutions.

A similar modeling idea was proposed by Adachi (2000, 2002). He proposed to model trends using homogeneity analysis (see Gifi (1990)), a variant of principal component analysis for nominal data. Such an approach renders a trend vector for every subject under study, which is smoothed using spline functions.

We advertised the trend vector model for cases where the number of measurements over subjects differs. We have to be careful, however, if these differences are due to drop-out or other forms of missing data. When data are missing in the GEE method they should be *missing completely at random* (Little and Rubin, 1987), that is, the missingness does not depend on the observed nor the missing data. If this is not the case, the GEE scheme can break down as is illustrated in Kenward et al. (1994). If this is the case, imputation strategies should be used to obtain complete series of observations. However, care should be exercised if much data are missing.

Acknowledgements

The author would like to thank an anonymous referee for his or her detailed suggestions for clarification and improvement of the paper. This research was conducted while the author was sponsored by the Netherlands Organisation for Scientific Research (NWO), Innovational Grant, no. 452-06-002.

References

Adachi, K., 2000. Optimal scaling of a longitudinal choice variable with time-varying representation of individuals. British Journal of Mathematical and Statistical Psychology 53, 233–253.
Adachi, K., 2002. Optimal quantification of a longitudinal indicator matrix: Homogeneity and smoothness analysis. Journal of Classification 19, 215–248.
Agresti, A., 2002. Categorical Data Analysis, 2nd ed. John Wiley and Sons, New York.
Crowder, M., 1995. On the use of a working correlation matrix in using generalized linear models for repeated measurements. Biometrika 82, 407–410.
De Rooij, M., 2008a. The analysis of change, newton's law of gravity and association model. Journal of the Royal Statistical Society, series A 171, 137–157.
De Rooij, M., 2008b. Ideal point discriminant analysis revisited with a special emphasis on visualization. Psychometrika (conditionally accepted paper).
De Rooij, M., Heiser, W.J., 2000. Triadic distance models for the analysis of asymmetric three-way proximity data. British Journal of Mathematical and Statistical Psychology 53, 99–119.
Diggle, P.J., Heagerty, P., Liang, K.-Y., Zeger, S.L., 2002. Analysis of Longitudinal Data. Oxford University Press, Oxford.
Gifi, A., 1990. Nonlinear Multivariate Analysis. John Wiley and Sons, Inc., Chichester.
Gower, J.C., Hand, D.J., 1996. Biplots. Chapman and Hall, London.
Hartzel, J., Agresti, A., Caffo, B., 2001. Multinomial logit random effects models. Statistical Modelling 1, 81–102.
Hedeker, D., Gibbons, R.D., 2006. Longitudinal Data Analysis. John Wiley and Sons, Inc., Hoboken, New Jersey.
Heiser, W.J., 1981. Unfolding analysis of proximity data. Ph.D. Thesis, Leiden University.
Kenward, M.G.E., Lesaffre, E., Molenberghs, G., 1994. An application of maximum likelihood and estimating equations to the analysis of ordinal data from a longitudinal study with cases missing at random. Biometrics 50, 945–953.
Kiers, H.A.L., Groenen, P.J.F., 2006. Visualizing dependence of bootstrap confidence intervals for methods yielding spatial configurations. In: Zani, S., Cerioli, A., Riani, M., Vichi, M. (Eds.), Data Analysis, Classification and the Forward Search. Springer, Berlin, pp. 119–126.
Liang, K.-Y., Zeger, S.L., 1986. Longitudinal data analysis using generalized linear models. Biometrika 73, 13–22.
Lipsitz, S.R., Kim, K., Zhao, L., 1994. Analysis of repeated categorical data using generalized estimating equations. Statistics in Medicine 13, 1149–1163.
Little, R.J., Rubin, D.B., 1987. Statistical Analysis with Missing Data. John Wiley and Sons, Inc., New York.
Lumley, T., 1996. Generalized estimating equations for ordinal data: A note on working correlation structures. Biometrics 52, 354–361.
Mathworks, 2006. MATLAB: The Language of Technical Computing. Mathworks, Natick.
Meulman, J., 1982. Homogeneity analysis of incomplete data. DSWO Press, Leiden.
Molenberghs, G., Verbeke, G., 2005. Models for Discrete Longitudinal Data. Springer, New York.
O'Hara Hines, R.J., 1997. Analysis of clustered polytomous data using generalized estimating equations and working covariance structures. Biometrics 53, 1552–1556.
Pan, W., Le, C.T., 2001. Bootstrap model selection in generalized linear models. Journal of Agricultural, Biological and Environmental Statistics 6, 49–61.
Pepe, M.S., Anderson, G.L., 1994. A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. Communications in Statistics. Simulation and computation 23, 939–951.
Sherman, M., le Cessie, S., 1997. A comparison between bootstrap methods and generalized estimating equations for correlated outcomes in generalized linear models. Communications in Statistics- Simulation and Computation 26, 901–925.
Sutradhar, B.C., Das, K., 1999. On the efficiency or regression estimators in generalised linear models for longitudinal data. Biometrika 86, 459–465.
Verbeke, G., Molenberghs, G., 2000. Linear Mixed Models for Longitudinal Data. Springer, New York.
Zielman, B., Heiser, W.J., 1993. The analysis of asymmetry by a slide-vector. Psychometrika 58, 101–114.