A multivariate extension of the dynamic logit model for longitudinal data based on a latent Markov heterogeneity structure*

Francesco Bartolucci† and Alessio Farcomeni‡

September 17, 2008

Abstract

For the analysis of multivariate categorical longitudinal data, we propose an extension of the dynamic logit model. The resulting model is based on a marginal parametrization of the conditional distribution of each vector of response variables given the covariates, the lagged response variables, and a set of subject-specific parameters for the unobserved heterogeneity. The latter ones are assumed to follow a first-order Markov chain. For the maximum likelihood estimation of the model parameters we outline an EM algorithm. The data analysis approach based on the proposed model is illustrated by a simulation study and an application to a dataset which derives from the Panel Study on Income Dynamics and concerns fertility and female participation to the labor market.

Keywords: EM algorithm; hidden Markov chains; marginal link function; panel data; state dependence.

*We are very grateful to the Editor, an Associate Editor and two anonymous Referees for helpful comments and suggestions.

†Dipartimento di Economia, Finanza e Statistica, Università di Perugia, 06123 Perugia, Italy, *e-mail:* bart@stat.unipg.it

‡Sapienza - Università di Roma, Piazzale Aldo Moro 5, 00185 Roma, Italy, *e-mail:* alessio.farcomeni@uniroma1.it

1 Introduction

Among the statistical and econometric models for binary longitudinal data, the *dynamic logic model* is of particular interest and finds application in many fields, such as in the study of the labor market (Hsiao, 2005). For each subject in the sample, this model assumes that the logit for the response variable at a given occasion depends on a set of strictly exogenous covariates, the lagged response variable, and a subject-specific parameter, which may be treated as a fixed or a random parameter. Given the presence of the lagged response variable among the regressors, the dynamic logit model may be considered as a transition model; see Molenberghs and Verbeke (2004). This lagged variable is included to capture the *state dependence* (Heckman, 1981b), i.e. the direct effect that experiencing a certain situation in the present has on the probability of experiencing the same situation in the future. This implies that the response variables for the same subject are not independent even conditionally on observable and unobservable covariates.

When the lagged response variable is omitted, the *static logit model* results. This model was extended to the case of bivariate binary longitudinal data by Ten Have and Morabia (1999), who relied on a *bivariate logistic transform* (Glonek and McCullagh, 1995) for this extension. A related model was proposed by Todem et al. (2007) for the analysis of multivariate ordinal longitudinal data. The latter is based on an ordinal probit link function and has a very flexible structure.

The subject-specific parameters, which are used in the dynamic logit model to take into account the unobserved heterogeneity between subjects, are assumed to be time-constant. This assumption is common to many other models for longitudinal data. However, if the effect of unobservable factors on the responses of a subject is not time-constant, there can be bias in the parameter estimates, in particular for the parameters of association between the response variables. In the econometric literature, this problem is usually overcome by relaxing the assumption of independence between the error terms used in the structural equations for the response variables at the different occasions; see Heckman (1981a) and Hyslop (1999).

In this paper, we propose a multivariate extension of the dynamic logit model in which the problem of adequately representing the unobservable heterogeneity is addressed by including a vector of subject-specific parameters which is time-varying and follows a first-order homogeneous Markov chain. To parameterize the conditional distribution of the vector of response variables, given the covariates, the lagged response variables, and the subject-specific parameters, we rely on a family of multivariate link functions formulated as in Colombi and Forcina (2001); this family has a structure similar to that of Glonek (1996) and is strongly related to the multivariate logistic transform of Glonek and McCullagh (1995). In fact, it is based on marginal logits for each response variable and marginal log-odds ratios for each pair of response variables, which may be of type *local*, *global* or *continuation* (Bartolucci et al., 2007a). Consequently, the proposed model may also be applied in the presence of more than two response variables which may also have more than two categories, whereas the models Ten Have and Morabia (1999) and Todem et al. (2007) are limited to bivariate data. Moreover, specific types of logit may be used with ordinal variables.

The proposed model also extends the latent Markov model of Wiggins (1973) in several directions and is related to the extension of the same model proposed by Vermunt et al. (1999). In fact, we also assume a latent Markov process, the states of which correspond to the different configurations of the subject-specific parameter vectors. The main difference is that in our approach the covariates have a direct effect on the response variables, whereas in the approach of Vermunt et al. (1999) these covariates have a direct effect on the initial and transition probabilities of the latent process; see also Bartolucci and Nigro (2007). Moreover, in our approach the response variables may be correlated even conditionally on the covariates and their dependence structure may be modeled in a meaningful way by exploiting the flexibility of the parametrization we adopt.

For the maximum likelihood estimation of the proposed model, we use an EM algorithm (Dempster et al., 1977). We derive ad-hoc recursions adapted from the hidden-Markov literature (MacDonald and Zucchini, 1997) for the efficient implementation of the E-step of this algorithm. We also deal with model selection and testing hypotheses on the parameters, such as the hypothesis that the transition matrix of the latent process is diagonal, so that the subject-specific vector of parameters is time-constant. Finally, we deal with prediction of the vector of responses and illustrate the Viterbi algorithm (Viterbi, 1967; Juang and Rabiner, 1991) for path prediction, i.e. prediction of the sequence of latent states of a given subject on the basis of his/her observable covariates and response variables. The approach based on the proposed model is illustrated by an application to a dataset coming from the Panel Study on Income Dynamics (PSID), which allows us to study the relation between fertility and woman participation to the labor market, a topic of great interest in labor economics (Hyslop, 1999; Carrasco, 2001).

The paper is organized as follows. In Section 2 we briefly review the relevant literature for our approach. In Section 3 we illustrate the proposed model for multivariate categorical longitudinal data. Likelihood inference for this model is outlined in Section 4. In Section 5 we show the results of a simulation study of the performance of the maximum likelihood estimator. The application to the PSID dataset is illustrated in Section 6. Final conclusions are drawn in Section 7.

The approach described in this paper has been implemented in a series of Matlab functions which are available from the JASA Supplemental Archive website.

2 Preliminaries

Let y_{it} denote the binary response variable for subject i at occasion t, with $i = 1,\dots,n$ and $t = 1,\dots,T$, and let $\boldsymbol{x}_{it}$ be the corresponding vector of strictly exogenous covariates. The dynamic logit model assumes that

$$\log\frac{p(y_{it}=1|\alpha_i,\boldsymbol{x}_{it},y_{i,t-1})}{p(y_{it}=0|\alpha_i,\boldsymbol{x}_{it},y_{i,t-1})} = \alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + y_{i,t-1}\gamma,$$

where α_i is a subject-specific parameter which captures the effect of unobservable covariates, $\boldsymbol{\beta}$ is a vector of regression coefficients for the observable covariates and γ is a parameter for the state dependence. Denoting by $1\{\cdot\}$ the indicator function, this model is justified in the econometric literature on the basis of the structural equations

$$y_{it} = 1\{\alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + y_{i,t-1}\gamma + \varepsilon_{it} > 0\}, \quad i = 1,\dots,n, \quad t = 1,\dots,T, \tag{1}$$

where ε_{it} are independent error terms with standard logistic distribution.

The subject-specific parameters may be treated as fixed or random. In the second case, the initial condition problem arises since the first available observation, y_{i0}, is correlated with the random parameter α_i. This correlation may be typically explained by considering that even this observation is generated from a distribution depending on observable and unobservable covariates which also affect the distribution of $y_{i1},\dots,y_{iT}$. For further details see Heckman (1981a) and Hsiao (2005, Sec. 7.5.2).

For the case in which we observe two binary response variables, denoted by y_{hit}, with $h = 1,2$, $i = 1,\dots,n$, $t = 1,\dots,T$, Ten Have and Morabia (1999) proposed a model which ignores state dependence and is based on the assumptions

$$\log\frac{p(y_{hit}=1|\alpha_{1i},\boldsymbol{x}_{it})}{p(y_{hit}=0|\alpha_{hi},\boldsymbol{x}_{it})} = \alpha_{hi} + \boldsymbol{x}_{it}'\boldsymbol{\beta}_h, \quad h = 1,2,$$

$$\log\frac{p(y_{1it}=1,y_{2it}=1|\alpha_{3i},\boldsymbol{x}_{it})p(y_{1it}=0,y_{2it}=0|\alpha_{3i},\boldsymbol{x}_{it})}{p(y_{1it}=1,y_{2it}=0|\alpha_{3i},\boldsymbol{x}_{it})p(y_{1it}=0,y_{2it}=1|\alpha_{3i},\boldsymbol{x}_{it})} = \alpha_{3i} + \boldsymbol{x}_{it}'\boldsymbol{\beta}_3.$$

The subject-specific parameters α_{1i}, α_{2i} and α_{3i} are assumed to be independent with standard normal distribution. This model also corresponds to a set of structural equations similar to (1) which involve bivariate error terms following a suitable copula of the standard logistic distribution.

The above models assume that the subject-specific parameters are time-invariant. This assumption may be relaxed by assuming that the error terms in structural equations of type (1) are serially correlated. A different strategy is here adopted which consists of assuming that the subject-specific parameters are time-varying and follow a Markov chain, so as to avoid any parametric assumption on their distribution.

3 Proposed model

Let r denote the number of categorical response variables observed at each occasion and denote by y_{hit} the hth response variable for subject i at occasion t, with $h = 1, \dots, r$, $i = 1, \dots, n$ and $t = 1, \dots, T$. This variable has l_h categories indexed from 0 to $l_h - 1$. Also let $\boldsymbol{y}_{it}$ denote the vector with elements y_{hit}, $h = 1, \dots, r$, and let $\boldsymbol{p}(\boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})$ denote the column vector for the conditional distribution of $\boldsymbol{y}_{it}$ given the covariates, the lagged response variables, and a vector $\boldsymbol{\alpha}_{it}$ of time-varying random effects. The entries of $\boldsymbol{p}(\boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})$ are the conditional probabilities $p(\boldsymbol{y}_{it}|\boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})$ for all the possible configurations of $\boldsymbol{y}_{it}$ arranged in lexicographical order. For example, with two response variables having respectively two and three categories, we have the configurations $(0, 0)$, $(0, 1)$, $(0, 2)$, $(1, 0)$, $(1, 1)$, $(1, 2)$.

The model we propose assumes that $\boldsymbol{y}_{it}$ is conditionally independent of $\boldsymbol{y}_{i0}, \dots, \boldsymbol{y}_{i,t-2}$, given $\boldsymbol{x}_{it}$, $\boldsymbol{y}_{i,t-1}$ and $\boldsymbol{\alpha}_{it}$, $t = 2, \dots, T$, and that the latent process $\boldsymbol{\alpha}_{i1}, \dots, \boldsymbol{\alpha}_{iT}$ follows a Markov chain with specific parameters. We now describe in detail the parametrizations adopted for the distribution of each response vector and for the latent process.

3.1 Distribution of the response variables

We rely on a family of multivariate link functions which allows us to directly model marginal (with respect to the other response variables) logits and log-odds ratios of type *local*, *global* or *continuation*. For the h-th variable, these logits are defined as follows for $z = 1, \dots, l_h - 1$:

- *local*: $$\log \frac{p(y_{hit} = z | \boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})}{p(y_{hit} = z - 1 | \boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})},$$
- *global*: $$\log \frac{p(y_{hit} \geq z | \boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})}{p(y_{hit} < z | \boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})},$$
- *continuation*: $$\log \frac{p(y_{hit} \geq z | \boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})}{p(y_{hit} = z - 1 | \boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})}.$$

Local logits are appropriate when the categories are not ordered. Logits of type global and continuation are suitable for ordinal variables. In particular, logits of type global are more appropriate when the variable may be seen as a discretized version of an underlying continuum, whereas logits of type continuation are more appropriate when its categories correspond to levels of achievement that may be entered only if the previous level has already been achieved.

Marginal log-odds ratios are defined as contrasts between conditional logits and their definition depends on the type of logit chosen for each response variable. For example, when local logits are used for variable h_1 and global logits for variable h_2, the following log-odds ratios result for $z_1 = 1, \dots, l_{h_1} - 1$ and $z_2 = 1, \dots, l_{h_2} - 1$:

$$\log \frac{p(y_{h_1it} = z_1, y_{h_2it} \geq z_2 | \boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1}) p(y_{h_1it} = z_1 - 1, y_{h_2it} < z_2 | \boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})}{p(y_{h_1it} = z_1 - 1, y_{h_2it} \geq z_2 | \boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1}) p(y_{h_1it} = z_1, y_{h_2it} < z_2 | \boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})}. \quad (2)$$

Once the type of logit has been chosen for each response variable, these logits and the corresponding log-odds ratios are collected in a vector which may be expressed as

$$\boldsymbol{\eta}(\boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1}) = \boldsymbol{C} \log[\boldsymbol{M} \boldsymbol{p}(\boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})], \quad (3)$$

where $\boldsymbol{C}$ and $\boldsymbol{M}$ are appropriate matrices whose construction is described in Colombi and Forcina (2001). In order to ensure that $\boldsymbol{\eta}(\boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})$ is a one-to-one function of $\boldsymbol{p}(\boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})$, we constrain to 0 all the three and higher-order log-linear interactions of the conditional distribution of $\boldsymbol{y}_{it}$, given $\boldsymbol{\alpha}_{it}$, $\boldsymbol{x}_{it}$ and $\boldsymbol{y}_{i,t-1}$. Invertibility of (3) then follows from Colombi and Forcina (2001) and to obtain $\boldsymbol{p}(\boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})$ from $\boldsymbol{\eta}(\boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})$ we can exploit the iterative algorithm they describe; see also Bartolucci et al. (2007a). MATLAB functions for constructing the matrices $\boldsymbol{C}$ and $\boldsymbol{M}$ in (3) and inverting this link function are available together with those for parameter estimation.

In order to relate the vector of marginal effects defined above to the covariates and the lagged response variables, we split it into the subvectors $\boldsymbol{\eta}_1(\boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})$ and $\boldsymbol{\eta}_2(\boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})$, which contain, respectively, marginal logits and log-odds ratios. We then assume that, for $i = 1, \ldots, n$ and $t = 1, \ldots, T$,

$$\boldsymbol{\eta}_1(\boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1}) = \boldsymbol{\alpha}_{it} + \boldsymbol{X}_{it}\boldsymbol{\beta} + \boldsymbol{Y}_{it}\boldsymbol{\gamma}, \tag{4}$$

$$\boldsymbol{\eta}_2(\boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1}) = \boldsymbol{\delta}, \tag{5}$$

where $\boldsymbol{X}_{it}$ and $\boldsymbol{Y}_{it}$ are suitable design matrices defined on the basis of, respectively, $\boldsymbol{x}_{it}$ and $\boldsymbol{y}_{i,t-1}$, whereas $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ are vectors of parameters.

As an example consider the case of $r = 3$ variables with two, three and three levels ($l_1 = 1$, $l_2 = 2$, $l_3 = 2$), which are treated with logits of type local, global and continuation, respectively. Overall, there are five logits which are expressed according to the above definition and eight log-odds ratios which are defined as in (2) for the first pair of response variables and in a similar way for the other two pairs. The logits may be parametrized as follows

$$\log \frac{p(y_{1it} = 1|\boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})}{p(y_{1it} = 0|\boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})} = \alpha_{1it} + \boldsymbol{x}'_{it}\boldsymbol{\beta}_1 + \boldsymbol{y}'_{i,t-1}\boldsymbol{\gamma}_1,$$

$$\log \frac{p(y_{2it} \geq z|\boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})}{p(y_{2it} < z|\boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})} = \alpha_{z+1,it} + \boldsymbol{x}_{it}'\boldsymbol{\beta}_2 + \boldsymbol{y}_{i,t-1}'\boldsymbol{\gamma}_2, \quad z = 1, 2, \tag{6}$$

$$\log \frac{p(y_{3it} \geq z|\boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})}{p(y_{3it} = z-1|\boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})} = \alpha_{z+3,it} + \boldsymbol{x}_{it}'\boldsymbol{\beta}_3 + \boldsymbol{y}_{i,t-1}'\boldsymbol{\gamma}_3, \quad z = 1, 2, \tag{7}$$

Note that, following a standard practice in marginal regression models for ordinal variables (see McCullagh (1980)), the regression coefficients for the covariates and those for the lagged response variables are the same for both logits in (6) and in (7). On the other hand, the intercepts α_{hit} are specific to each response category.

The above parametrization may be casted into (4) with

$$\boldsymbol{X}_{it} = \begin{pmatrix} \boldsymbol{x}_{it}' & \boldsymbol{0}' & \boldsymbol{0}' \\ \boldsymbol{0}' & \boldsymbol{x}_{it}' & \boldsymbol{0}' \\ \boldsymbol{0}' & \boldsymbol{x}_{it}' & \boldsymbol{0}' \\ \boldsymbol{0}' & \boldsymbol{0}' & \boldsymbol{x}_{it}' \\ \boldsymbol{0}' & \boldsymbol{0}' & \boldsymbol{x}_{it}' \end{pmatrix}, \quad \boldsymbol{Y}_{it} = \begin{pmatrix} \boldsymbol{y}_{i,t-1}' & \boldsymbol{0}' & \boldsymbol{0}' \\ \boldsymbol{0}' & \boldsymbol{y}_{i,t-1}' & \boldsymbol{0}' \\ \boldsymbol{0}' & \boldsymbol{y}_{i,t-1}' & \boldsymbol{0}' \\ \boldsymbol{0}' & \boldsymbol{0}' & \boldsymbol{y}_{i,t-1}' \\ \boldsymbol{0}' & \boldsymbol{0}' & \boldsymbol{y}_{i,t-1}' \end{pmatrix}, \tag{8}$$

where $\boldsymbol{0}$ denotes a column vector of zeros of suitable dimension, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \boldsymbol{\beta}_3')'$, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1', \boldsymbol{\gamma}_2', \boldsymbol{\gamma}_3')'$ and $\boldsymbol{\alpha}_{it}$ being a vector with elements $\alpha_{1it}, \ldots, \alpha_{5it}$. Finally, because of assumption (5), each log-odds ratio is simply equal to a specific element of $\boldsymbol{\delta}$.

3.2 Distribution of the subject-specific parameters

For each subject i, the random parameter vectors $\boldsymbol{\alpha}_{it}$, $t = 1, \ldots, T$, are assumed to follow a first-order Markov chain with states $\boldsymbol{\xi}_c$, for $c = 1, \ldots, k$, and initial probabilities $\lambda_c(\boldsymbol{y}_{i0}) = p(\boldsymbol{\alpha}_{i1} = \boldsymbol{\xi}_c|\boldsymbol{y}_{i0})$ collected in the column vector $\boldsymbol{\lambda}(\boldsymbol{y}_{i0})$. The transition probabilities are denoted by $\pi_{cd} = p(\boldsymbol{\alpha}_{it} = \boldsymbol{\xi}_d|\boldsymbol{\alpha}_{i,t-1} = \boldsymbol{\xi}_c)$, $c, d = 1, \ldots, k$, $t = 2, \ldots, T$, and are collected in the matrix $\boldsymbol{\Pi}$.

In order to take the initial condition problem into account (see Section 2), the probabilities $\lambda_c(\boldsymbol{y}_{i0})$ are allowed to depend on the initial observation. In particular, let $\boldsymbol{\psi}(\boldsymbol{y}_{i0})$ be the $(k-1)$-dimensional column vector of the logits $\log[\lambda_c(\boldsymbol{y}_{i0})/\lambda_1(\boldsymbol{y}_{i0})]$, $c = 2, \ldots, k$. We assume that

$$\boldsymbol{\psi}(\boldsymbol{y}_{i0}) = \boldsymbol{Y}_{i0}\boldsymbol{\phi}, \tag{9}$$

where $\boldsymbol{Y}_{i0}$ is a design matrix depending on $\boldsymbol{y}_{i0}$ and $\boldsymbol{\phi}$ is the corresponding vector of parameters. Typically, this matrix is equal to $\boldsymbol{I}_{k-1} \otimes (\,1 \quad \boldsymbol{y}'_{i0}\,)$, with $\boldsymbol{I}_z$ denoting an identity matrix of dimension z.

Note that, by assumption, the initial and transition probabilities of the latent process are independent of the covariates. This assumption could be easily relaxed by adopting a parametrization similar to that used by Vermunt et al. (1999). However, we prefer to retain this assumption so that the effect of the covariates and that of the state dependence are entirely captured by the parameters in $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ through (4).

Finally, consider that assuming a discrete rather than a continuous latent process avoids the need of parametric assumptions and simplifies the estimation of the resulting model from the computational point of view. In fact, as we show in the following section, the likelihood of the model can be exactly computed without the need of quadrature or Monte Carlo methods, which would be required if the latent process was assumed to be continuous. From the computational point of view it could be objected that the number of elements of the transition matrix increases with the square of the number of latent states. However, if necessary, the model may be made more parsimonious by imposing a specific structure for this matrix. For instance, we can require that all the off-diagonal elements are equal each other or that this matrix is symmetric; see Bartolucci (2006) for examples of this type.

On the other hand, the assumption that the process representing the evolution of a latent characteristic is discrete rather than continuous may not be realistic in certain situations. Our hope is that in most of these situations the discrete process adequately approximates the continuous process and then our model gives a realistic representation of the data generation mechanism, especially when a large number of states is adopted and the continuous process has a Markovian dependence structure, e.g. AR(1). This is in agreement with the practice commonly adopted in the latent variable literature of assuming a discrete distribution for a latent trait which has a continuous nature; see, for instance, Lindsay et al. (1991). However, with reference to our context, theoretical results on the quality of the approximation and on the implications on the parameter estimation are not available and then in Section 5 we provide some results based on simulation.

4 Likelihood inference

Inference for the proposed model is based on the log-likelihood

$$\ell(\boldsymbol{\theta}) = \sum_i \log[p(\boldsymbol{y}_{i1}, \ldots, \boldsymbol{y}_{iT} | \boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iT}, \boldsymbol{y}_{i0})],$$

where $\boldsymbol{\theta}$ is a short hand notation for all the non-redundant model parameters corresponding to the vectors $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\delta}$ and $\boldsymbol{\phi}$ and the off-diagonal elements of the matrix $\boldsymbol{\Pi}$. The model assumptions imply that $p(\boldsymbol{y}_{i1}, \ldots, \boldsymbol{y}_{iT} | \boldsymbol{x}_{i1} \ldots, \boldsymbol{x}_{iT}, \boldsymbol{y}_{i0})$ is equal to

$$\sum_{\boldsymbol{\alpha}_{i1}} \cdots \sum_{\boldsymbol{\alpha}_{iT}} \left[p(\boldsymbol{\alpha}_{i1} | \boldsymbol{y}_{i0}) \prod_{t>1} p(\boldsymbol{\alpha}_{it} | \boldsymbol{\alpha}_{i,t-1}) \prod_t p(\boldsymbol{y}_{it} | \boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1}) \right], \qquad (10)$$

with the sum $\sum_{\boldsymbol{\alpha}_{it}}$ extended to all the possible configurations of $\boldsymbol{\alpha}_{it}$. An efficient rule to compute the probability in (10) is given in Appendix.

4.1 Estimation

In order to estimate $\boldsymbol{\theta}$, we maximize $\ell(\boldsymbol{\theta})$ by using a version of the EM algorithm (Dempster et al., 1977) which may be implemented along the same lines as in Bartolucci (2006) and Bartolucci et al. (2007b). However, these papers deal with versions of the latent Markov model which are based on a much simpler parametrization of the conditional distribution of the response variables and include categorical covariates only.

The EM algorithm alternates the following steps until convergence:

- **E-step**: compute the conditional expected value of the complete data log-likelihood given the observed data and $\tilde{\boldsymbol{\theta}}$, the current estimate of $\boldsymbol{\theta}$;

- **M-step**: maximize the expected value above with respect to $\boldsymbol{\theta}$.

Let w_{itc} denote a dummy variable equal to 1 if subject i is in latent state c at occasion t (i.e. $\boldsymbol{\alpha}_{it} = \boldsymbol{\xi}_c$) and to 0 otherwise. The *complete data log-likelihood*, that we could compute if we knew these dummy variables at every occasion, is:

$$\ell^*(\boldsymbol{\theta}) = \sum_i \sum_c \Big\{ \sum_t w_{itc} \log[p(\boldsymbol{y}_{it}|\boldsymbol{\alpha}_{it} = \boldsymbol{\xi}_c, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})] + w_{i1c} \log[\lambda_c(\boldsymbol{y}_{i0})] + \sum_d z_{icd} \log(\pi_{cd}) \Big\},$$

where $z_{icd} = \sum_{t>1} w_{i,t-1,c} w_{itd}$ is equal to the number of times subject i moves from state c to state d. The conditional expected value of $\ell^*(\boldsymbol{\theta})$ at the E-step is then given by the same expression as above in which we substitute the variables w_{itc} and z_{icd} with the corresponding expected values. These are equal to

$$\tilde{w}_{itc}(\tilde{\boldsymbol{\theta}}) = p(\boldsymbol{\alpha}_{it} = \boldsymbol{\xi}_c | \boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iT}, \boldsymbol{y}_{i0}, \ldots, \boldsymbol{y}_{iT}), \quad (11)$$

$$\tilde{z}_{icd}(\tilde{\boldsymbol{\theta}}) = \sum_{t>1} p(\boldsymbol{\alpha}_{i,t-1} = \boldsymbol{\xi}_c, \boldsymbol{\alpha}_{it} = \boldsymbol{\xi}_d | \boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iT}, \boldsymbol{y}_{i0}, \ldots, \boldsymbol{y}_{iT}), \quad (12)$$

with the *posterior probabilities* in (11) and (12) evaluated at $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$. Efficient computation of these probabilities may be carried out as described in Appendix. The conditional expected value of $\ell^*(\boldsymbol{\theta})$ is denoted by $\tilde{\ell}^*(\boldsymbol{\theta}|\tilde{\boldsymbol{\theta}})$.

At the M-step, $\tilde{\ell}^*(\boldsymbol{\theta}|\tilde{\boldsymbol{\theta}})$ is maximized by separately maximizing its components:

$$\tilde{\ell}_1^*(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}|\tilde{\boldsymbol{\theta}}) = \sum_i \sum_c \sum_t \tilde{w}_{itc}(\tilde{\boldsymbol{\theta}}) \log[p(\boldsymbol{y}_{it}|\boldsymbol{\alpha}_{it} = \boldsymbol{\xi}_c, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})],$$

$$\tilde{\ell}_2^*(\boldsymbol{\phi}|\tilde{\boldsymbol{\theta}}) = \sum_i \sum_c \tilde{w}_{i1c}(\tilde{\boldsymbol{\theta}}) \log[\lambda_c(\boldsymbol{y}_{i0})],$$
$$\tilde{\ell}_3^*(\boldsymbol{\Pi}|\tilde{\boldsymbol{\theta}}) = \sum_i \sum_c \sum_d \tilde{z}_{icd}(\tilde{\boldsymbol{\theta}}) \log(\pi_{cd}).$$

An explicit solution is available to maximize the last one, which consists of letting each π_{cd} proportional to $\sum_i \tilde{z}_{icd}(\tilde{\boldsymbol{\theta}})$ for $c, d = 1, \ldots, k$. To maximize $\tilde{\ell}_2^*(\boldsymbol{\phi}|\tilde{\boldsymbol{\theta}})$ we can use a standard iterative algorithm of Newton-Raphson type for multinomial logit models. A Newton-Raphson algorithm may also be used to maximize $\tilde{\ell}_1^*(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}|\tilde{\boldsymbol{\theta}})$. This algorithm is slightly more complex than that for maximizing $\tilde{\ell}_2^*(\boldsymbol{\phi}|\tilde{\boldsymbol{\theta}})$ since, at each step, it requires inversion of (3) for every i and t and the k possible values of $\boldsymbol{\alpha}_{it}$; details on its implementation may be deduced from Colombi and Forcina (2001).

We take the value of $\boldsymbol{\theta}$ at convergence of the EM algorithm as the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$. As is typical for latent variable models, the likelihood may be multimodal and the point at convergence depends on the starting values for the parameters, which then need to be carefully chosen. At this regard, we follow a rule which consists of a preliminary fitting of a model based on assumptions (4) and (5) under the constraint $\boldsymbol{\alpha}_{it} = \bar{\boldsymbol{\xi}}$, $i = 1, \ldots, n$, $t = 1, \ldots, T$. This is a simplified version of our model which, being based on a common intercept $\bar{\boldsymbol{\xi}}$ for all subjects and occasions, rules out unobserved heterogeneity. In this way we directly obtain the initial values for $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$, whereas, for $c = 1, \ldots, k$, the initial value of $\boldsymbol{\xi}_c$ is found by adding a suitable constant f_c to each element of the estimate of $\bar{\boldsymbol{\xi}}$. Finally, we use $\mathbf{0}$ as starting value for $\boldsymbol{\phi}$ and, for a suitable constant s, $(\mathbf{1}_k\mathbf{1}_k' + s\boldsymbol{I}_k)/(k+s)$ as starting value for $\boldsymbol{\Pi}$, where $\mathbf{1}_h$ denotes a column vector of h ones. In our implementation, we choose $f_1, \ldots, f_k$ as k equispaced points from -2.5 to 2.5 and we let $s = 9$. To check that the EM algorithm converges to the global maximum of the likelihood, we also suggest to try different starting values for the parameters which may be generated by randomly perturbating those obtained by the deterministic rule above. For instance, a random number with normal distribution with zero mean may be added to the initial value of each element of $\boldsymbol{\xi}_c$, to that of $\boldsymbol{\beta}$ and so on.

On the basis of some experiments based on simulated data and on the PSID dataset illustrated in Section 6, we can conclude that the chance that the likelihood is multimodal grows as the number of latent states increases and as the sample size decreases. Moreover, imposing a suitable constraint on the transition matrix $\mathbf{\Pi}$ considerably reduces the chance that the likelihood is multimodal. In particular, for the PSID dataset we observed that the likelihood of the unrestricted model has a few local maxima with $k \geq 3$ latent states. In any case, these local maxima may be easily found by the random initialization mechanism for the EM algorithm outlined above and their number dramatically reduces under the constraint that the off-diagonal elements of the transition matrix are equal each other. Moreover, the best solution usually corresponds to that found starting with the deterministic rule.

A final point concerns how to compute the information matrix. For this aim, several methods have been proposed in the literature which exploit the results of the EM algorithm; see McLachlan and Peel (2000, Ch. 2) and the references therein. In our context, these methods can not be directly applied, so we prefer to obtain the observed information matrix, denoted by $\boldsymbol{J}(\boldsymbol{\theta})$, as minus the numerical derivative of the score vector $\boldsymbol{s}(\boldsymbol{\theta})$ which corresponds to the first derivative of $\tilde{\ell}(\boldsymbol{\theta}|\tilde{\boldsymbol{\theta}})$ with respect to $\boldsymbol{\theta}$, evaluated at $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta}$. The latter is already used at the M-step and then computation of the observed information matrix requires a small extra code to be implemented. The observed information matrix at the maximum likelihood estimate, $\boldsymbol{J}(\hat{\boldsymbol{\theta}})$, may be used to check local identifiability of the model and to compute the standard errors $se(\hat{\boldsymbol{\theta}})$ in the usual way. The validity of this procedure to obtain standard errors for $\hat{\boldsymbol{\theta}}$ is assessed by simulation at the end of Section 5.

4.2 Model selection and hypotheses testing

A fundamental problem is that of the choice of the number of latent states, denoted by k. In the literature on latent variable models and finite mixture models, see in particular McLachlan and Peel (2000, Ch. 6), the most used criteria are the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). According to these criteria, we choose the number of states corresponding to the minimum of $AIC = -2\ell(\hat{\boldsymbol{\theta}}) + 2g$ and $BIC = -2\ell(\hat{\boldsymbol{\theta}}) + g\log(n)$, respectively. These two indices involve penalization terms depending on g, the number of non-redundant parameters, which is equal to the sum of:

- the number of columns of each design matrix $\boldsymbol{X}_{it}$ in (4) which is at most equal to $\sum_h(l_h - 1)$ times the number of covariates, where $\sum_h(l_h - 1)$ is the number of marginal logits;
- the number of columns of the matrix $\boldsymbol{Y}_{it}$ in (4), at most $r\sum_h(l_h - 1)$;
- $\sum_{h1<r}\sum_{h2>h_1}(l_{h_1} - 1)(l_{h_2} - 1)$ which corresponds to the number of marginal log-odds ratio and then to the dimension of $\boldsymbol{\delta}$;
- $k\sum_h(l_h - 1)$ corresponding to the number of elements of the vectors $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_k$;
- the number of columns of the design matrix $\boldsymbol{Y}_{i0}$ in (9), typically $r(k-1)$;
- $k(k-1)$ which corresponds to the number of independent transition probabilities collected in $\boldsymbol{\Pi}$.

Given the different penalization terms involved in the two indices above, the two criteria do not always lead to choosing the same number of latent states. Some suggestions on their use are given in Section 5, where these are studied by simulation.

Once the number of latent states has been chosen, it may be interesting to test hypotheses on the parameters. Under the usual regularity conditions, these hypotheses may be tested by using Wald statistics based on the standard errors computed as mentioned above. This is convenient when the hypothesis of interest is that one of the parameters in $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ or $\boldsymbol{\delta}$ is equal to 0. A more general method to test hypotheses is based on the likelihood ratio statistic $D = -2[\ell(\hat{\boldsymbol{\theta}}_0) - \ell(\hat{\boldsymbol{\theta}})]$, where $\hat{\boldsymbol{\theta}}_0$ is the maximum likelihood estimate of $\boldsymbol{\theta}$ under the hypothesis of interest, which may be computed by the same EM algorithm illustrated in Section 4.1. Under standard regularity conditions, a p-value for this statistic can be computed on the basis of a chi-squared distribution with the appropriate number of degrees of freedom.

A hypothesis of particular interest is that the transition matrix is diagonal. Rejecting this hypothesis implies that the effect of unobserved factors on the response variables is not time-constant so that conventional models, such as the dynamic logit model, are not suitable for the data at hand. To test this hypothesis we can use the likelihood ratio statistic defined above, but a boundary problem occurs, since it corresponds to the constraint that all the off-diagonal transition probabilities are equal to zero. Then the approximation of the likelihood ratio null distribution by a chi-squared distribution is not valid anymore. We can instead use the result of Bartolucci (2006), who showed that the likelihood ratio statistic for hypotheses on the transition matrix of a latent Markov model has null asymptotic distribution of chi-bar-squared type, i.e. a mixture of chi-squared distributions (Shapiro, 1988; Silvapulle and Sen, 2004). This implies that the p-value for an observed value d of D may be computed as

$$\Pr(D > d) = \sum_{h=0}^{k(k-1)} w_h \Pr(C_h > d),$$

where C_h has chi-squared distribution with h degrees of freedom and the weights w_h can be computed through a simple Monte Carlo procedure. This procedure consists of drawing a large number of parameter vectors from the asymptotic distribution of the unconstrained maximum likelihood estimator and computing the proportion of vectors which violate the constraint of interest; see also Dardanoni and Forcina

(1998).

Finally, note that a likelihood ratio test statistic may also be used to choose the number of latent states by comparing the model with k and that with $k+1$ states for increasing values of k. However, the significance of this statistic needs to be valuated by a bootstrap procedure; we prefer to avoid this selection criterion because too computationally intensive.

4.3 Prediction of the response vector and path prediction

Once the model has been fitted, it is usually of interest to predict the response vector for subject i at occasion t on the basis of the vector of covariates $\boldsymbol{x}_{it}$ and the lagged response vector $\boldsymbol{y}_{i,t-1}$. A natural way to predict this response vector, denoted by $\hat{\boldsymbol{y}}_{it}$, is by maximizing with respect to $\boldsymbol{y}$ the manifest probability

$$p(\boldsymbol{y}_{it}=\boldsymbol{y}|\boldsymbol{x}_{it},\boldsymbol{y}_{i,t-1})=\sum_c p(\boldsymbol{y}_{it}=\boldsymbol{y}|\boldsymbol{\alpha}_{it}=\boldsymbol{\xi}_c,\boldsymbol{x}_{it},\boldsymbol{y}_{i,t-1})p(\boldsymbol{\alpha}_{it}=\boldsymbol{\xi}_c|\boldsymbol{y}_{i0}),$$

once it has been computed on the basis of the maximum likelihood estimate of $\boldsymbol{\theta}$.

Another problem of interest is that of predicting the state $\hat{c}_{it}$ of subject i at a given time occasion t. The estimate is the maximal-a-posteriori prediction based on the probabilities in (11), which are obtained as a by-result of the EM algorithm.

A related problem is that of predicting the entire sequence of latent states for subject i, which corresponds to the maximum with respect to $c_1,\ldots,c_T$ of the posterior probability $p(\boldsymbol{\alpha}_{i1}=\boldsymbol{\xi}_{c_1},\ldots,\boldsymbol{\alpha}_{iT}=\boldsymbol{\xi}_{c_T}|\boldsymbol{x}_{i1},\ldots,\boldsymbol{x}_{iT},\boldsymbol{y}_{i0},\ldots,\boldsymbol{y}_{iT})$. The predicted path is denoted by $\tilde{c}_{i1},\ldots,\tilde{c}_{iT}$ and it is not ensured to be equal to $\hat{c}_{i1},\ldots,\hat{c}_{iT}$, when each $\hat{c}_{it}$ is found as above on the basis of the posterior probabilities in (11). In particular, the previous method does not take into account the joint probability of the latent sequence, and may even produce inconsistent sequences.

To predict the entire sequence of latent states we can use the Viterbi algorithm

(Viterbi, 1967; Juang and Rabiner, 1991). Let $\rho_{i1}(c) = p(\boldsymbol{\alpha}_{i1} = \boldsymbol{\xi}_c, \boldsymbol{y}_{i1}|\boldsymbol{x}_{i1}, \boldsymbol{y}_{i0})$ and, for $t = 2, \ldots, T$, let

$$\rho_{it}(c) = \max_{c_1,\ldots,c_{t-1}} p(\boldsymbol{\alpha}_{i1} = \boldsymbol{\xi}_{c_1}, \ldots, \boldsymbol{\alpha}_{i,t-1} = \boldsymbol{\xi}_{c_{t-1}}, \boldsymbol{\alpha}_{it} = \boldsymbol{\xi}_c, \boldsymbol{y}_{i1}, \ldots, \boldsymbol{y}_{it}|\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{it}, \boldsymbol{y}_{i0}).$$

The algorithm performs a forward recursion in order to compute the above quantities, and then it finds the most likely latent sequence with a backward recursion.

More precisely, the algorithm performs the following steps:

1. for $i = 1, \ldots, n$ and $c = 1, \ldots, k$ compute $\rho_{i1}(c)$ as $\lambda_c(\boldsymbol{y}_{i0})p(\boldsymbol{y}_{i1}|\boldsymbol{\alpha}_{i1} = \boldsymbol{\xi}_c, \boldsymbol{x}_{i1})$;
2. for $i = 1, \ldots, n$, $t = 2, \ldots, T$ and $d = 1, \ldots, k$ compute $\rho_{it}(d)$ as
$$p(\boldsymbol{y}_{i,t+1}|\boldsymbol{\alpha}_{i,t+1} = \boldsymbol{\xi}_d, \boldsymbol{x}_{i,t+1}) \ \max_c[\rho_{i,t-1}(c)\pi_{cd}];$$
3. for $i = 1, \ldots, n$ find the optimal state $\tilde{c}_{iT}$ as $\tilde{c}_{iT} = \arg\max_c \rho_{iT}(c)$;
4. for $i = 1, \ldots, n$ and $t = T-1, \ldots, 1$ find $\tilde{c}_{it}$ as $\tilde{c}_{it} = \arg\max_c \rho_{it}(c)\pi_{c,\tilde{c}_{i,t+1}}$.

All the above quantities are computed on the basis of the maximum likelihood estimate of the parameter $\boldsymbol{\theta}$ of the model of interest.

5 Simulation study

In order to assess the properties the maximum likelihood estimator described in Section 4.1, we performed a simulation study which is described below. The same study allows us to assess the performance of the selection criteria described in Section 4.2.

5.1 Simulation design

We considered two scenarios, the first with two response variables (both binary) and the second with three response variables (the first with two and the others with three categories). Under each scenario, we considered two continuous covariates and generated 1000 samples from the proposed model with $T = 4, 8$ (panel length), $n = 500, 1000$ (sample size) and $k = 1, 2, 3$ (number of latent states). For each sample we computed the maximum likelihood estimate of the parameters under the assumed model and the corresponding standard errors. We also predicted the optimal number of states according to the AIC and BIC criteria. In order to verify the effect of model misspecification, we considered a further setting in which the subject-specific parameters follow a continuous process.

With $r = 2$ response variables, the design matrices in (4) are defined as $\boldsymbol{X}_{it} = \boldsymbol{1}_2 \otimes \boldsymbol{x}_{it}'$ where the two covariates in $\boldsymbol{x}_{it}$ are independently generated from a standard normal distribution for $i = 1, \dots, n$ and $t = 1, \dots, T$. Moreover, $\boldsymbol{Y}_{it} = \boldsymbol{1}_2 \otimes \boldsymbol{y}_{i,t-1}'$ and, for $k \geq 2$, the design matrix $\boldsymbol{Y}_{i0}$ in (9) is defined as $\boldsymbol{1}_{k-1} \otimes (\, 1 \quad \boldsymbol{y}_{i0}' \,)$, where the initial observations in $\boldsymbol{y}_{i0}$ are independently generated from a Bernoulli distribution with parameter 0.5 for $i = 1, \dots, n$. The true values of the regression parameters are chosen as $\boldsymbol{\beta} = (\, 1 \quad -1 \quad 1 \quad -1 \,)'$ and those of the parameters for the lagged responses are chosen as $\boldsymbol{\gamma} = (\, 1 \quad -1 \quad -1 \quad 1 \,)'$; we also let $\delta = -1$. According to the value of k, the parameters for the latent process are chosen as follows:

- $k = 1$: $\boldsymbol{\xi}_1 = (\, 0 \quad 0 \,)'$, $\lambda_1(\boldsymbol{y}_{i0}) = 1$, $\pi_{11} = 1$ (the latent process is degenerate);
- $k = 2$: $\boldsymbol{\xi}_1 = (\, -1 \quad -1 \,)'$ and $\boldsymbol{\xi}_2 = -\boldsymbol{\xi}_1$, with $\boldsymbol{\phi} = \boldsymbol{0}$ and transition matrix

$$\boldsymbol{\Pi} = \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}; \tag{13}$$

- $k = 3$: $\boldsymbol{\xi}_1 = (\, -2.5 \quad -2.5 \,)'$, $\boldsymbol{\xi}_2 = (\, 0 \quad 0 \,)'$ and $\boldsymbol{\xi}_3 = -\boldsymbol{\xi}_1$, with $\boldsymbol{\phi} = \boldsymbol{0}$ and

$$\boldsymbol{\Pi} = \begin{pmatrix} 0.80 & 0.15 & 0.05 \\ 0.10 & 0.80 & 0.10 \\ 0.05 & 0.15 & 0.80 \end{pmatrix}. \tag{14}$$

For $r = 3$ response variables we adopted the same parametrization described in the example in Section 3.1, which is based on local logits for the first variable (having two levels), global logits for the second variable (having three levels) and continuation logits for the third (having three levels) and on the design matrices defined in (8). For what concerns the parametrization of the latent process, we let $\boldsymbol{Y}_{i0} = \boldsymbol{1}_{k-1} \otimes (\,1 \quad \boldsymbol{1}_r'\boldsymbol{y}_{i0}/r\,)$ for $k \geq 2$, where the initial observations in $\boldsymbol{y}_{i0}$ are randomly generated from uniform discrete distributions with suitable support. We also let $\boldsymbol{\beta} = (\,1 \quad -1 \quad 1 \quad -1 \quad -1 \quad 1\,)'$, $\boldsymbol{\gamma} = (\,1 \quad 1 \quad -1\,)'$ and $\boldsymbol{\delta} = (\,1 \quad 1 \quad 0 \quad 0 \quad -1 \quad -1 \quad -1 \quad -1\,)'$. Note that the first two elements of $\boldsymbol{\delta}$ refer to the log-odds ratios for the pair of response variables (y_{1it}, y_{2it}), the second two refer to the log-odd ratios for (y_{1it}, y_{3it}), and the remaining ones refer to the log-odds ratio for (y_{2it}, y_{3it}). Moreover, for what concerns the parametrization of the latent process, with $k = 1$ we assumed $\boldsymbol{\xi}_1 = \bar{\boldsymbol{\xi}}$, where $\bar{\boldsymbol{\xi}} = (\,0 \quad 1 \quad -1 \quad 1 \quad -1\,)'$. With $k = 2$ we assumed $\boldsymbol{\xi}_1 = \bar{\boldsymbol{\xi}} - \boldsymbol{1}_2$ and $\boldsymbol{\xi}_2 = \bar{\boldsymbol{\xi}} + \boldsymbol{1}_2$; we also let $\boldsymbol{\phi} = \boldsymbol{0}$, with $\boldsymbol{\Pi}$ defined as in (13). Finally, with $k = 3$ we assumed $\boldsymbol{\xi}_1 = \bar{\boldsymbol{\xi}} - 2.5 \cdot \boldsymbol{1}_2$, $\boldsymbol{\xi}_2 = \bar{\boldsymbol{\xi}}$ and $\boldsymbol{\xi}_3 = \bar{\boldsymbol{\xi}} + 2.5 \cdot \boldsymbol{1}_2$ and that $\boldsymbol{\phi} = \boldsymbol{0}$, with $\boldsymbol{\Pi}$ as in (14).

The simulation settings in which the subject specific parameters follow a continuous process were formulated as above for both $r = 2$ and $r = 3$. The only difference is that $\boldsymbol{\alpha}_{it} = \boldsymbol{\varepsilon}_{it}$ when $r = 2$ and $\boldsymbol{\alpha}_{it} = \bar{\boldsymbol{\xi}} + \boldsymbol{\varepsilon}_{it}$ when $r = 3$, where, for $i = 1, \ldots, n$ and $t = 1, \ldots, T$, each element of $\boldsymbol{\varepsilon}_{it}$ is independently generated from an AR(1) process with correlation coefficient 0.9 and marginal variance equal to 2.

5.2 Simulation Results

For $r = 2$, the simulation results in terms of bias and standard deviation of the maximum likelihood estimator of each parameter of interest are shown in Table 1 (when $k = 2$) and in Table 2 (when $k = 3$), together with the average and the interquartile range of the standard errors computed for every sample. In both tables, $\hat{\beta}_h$ and $\hat{\gamma}_h$ denote, respectively, the h-th element of the estimator $\hat{\boldsymbol{\beta}}$ and that of the estimator $\hat{\boldsymbol{\gamma}}$, whereas $\hat{\alpha}_h$ denotes h-th element of the weighted mean of the vectors $\hat{\boldsymbol{\xi}}_1, \ldots \hat{\boldsymbol{\xi}}_k$, with weights equal to the posterior probability of each state. Each $\hat{\alpha}_h$ is an estimator of the average effect of the unobservable covariates on the corresponding marginal logit in (4).

We can observe that, with both $k = 2$ and $k = 3$, the bias of each estimator is always moderate and decreases as n and T increase. Moreover, the standard deviation decreases at the expected rate of $\sqrt{n}$ with respect to n and at a faster rate with respect to T. Obviously, the standard deviation is higher with $k = 3$ than with $k = 2$. Finally, for each estimator, the average standard error is always very close to the standard deviation; these standard errors have also a very low variability from sample to sample.

In order to evaluate the performance of AIC and BIC as selection criteria for the number of latent states, in Table 3 we report the frequency distribution of the predicted k under each simulation setting considered with $r = 2$. We can observe that AIC performs considerably well in all cases. In fact, the predicted k is only occasionally different from the true one and, when this happens, the former is always larger than the latter. On the other hand, BIC has an excellent behavior with the exception of the case $T = 4$, $n = 500$ and $k = 3$ when it tends to predict $k = 2$. As may be expected, this criterion performs better as the amount of information in the data increases. In fact, for the cases in which $T = 8$, BIC always singled out the true number of latent states.

With $r = 3$ response variables we obtained results similar to those commented above for $r = 2$ in terms of performance of the maximum likelihood estimator and the AIC and BIC selection criteria for the number of states. Some of these results

Est.	$T=4,\ n=500$ Bias	s.d.	ave. s.e.	IQR s.e.	$T=4,\ n=1000$ Bias	s.d.	ave. s.e.	IQR s.e.
$\hat{\alpha}_1$	0.004	-	-	-	0.001	-	-	-
$\hat{\beta}_1$	0.012	0.084	0.084	0.012	0.006	0.056	0.057	0.006
$\hat{\beta}_2$	-0.012	0.084	0.083	0.012	-0.007	0.057	0.056	0.006
$\hat{\gamma}_1$	-0.008	0.140	0.139	0.015	0.002	0.104	0.103	0.007
$\hat{\gamma}_2$	-0.002	0.149	0.148	0.013	0.001	0.102	0.101	0.006
$\hat{\alpha}_2$	0.001	-	-	-	0.001	-	-	
$\hat{\beta}_3$	-0.012	0.084	0.084	0.012	-0.006	0.58	0.057	0.005
$\hat{\beta}_4$	0.008	0.084	0.082	0.012	0.002	0.057	0.058	0.006
$\hat{\gamma}_3$	0.003	0.142	0.141	0.013	0.004	0.104	0.104	0.006
$\hat{\gamma}_4$	-0.006	0.150	0.149	0.014	-0.007	0.102	0.099	0.007
$\hat{\delta}$	-0.043	0.270	0.266	0.045	-0.019	0.175	0.177	0.022

Est.	$T=8,\ n=500$ Bias	s.d.	ave. s.e.	IQR s.e.	$T=8,\ n=1000$ Bias	s.d.	ave. s.e.	IQR s.e.
$\hat{\alpha}_1$	0.005	-	-	-	0.001	-	-	-
$\hat{\beta}_1$	0.007	0.053	0.054	0.004	0.003	0.038	0.038	0.002
$\hat{\beta}_2$	-0.006	0.055	0.054	0.004	-0.003	0.039	0.038	0.002
$\hat{\gamma}_1$	-0.006	0.099	0.099	0.005	-0.001	0.068	0.068	0.003
$\hat{\gamma}_2$	-0.005	0.102	0.101	0.005	-0.001	0.071	0.071	0.002
$\hat{\alpha}_2$	0.001	-	-	-	0.002	-	-	-
$\hat{\beta}_3$	0.001	0.055	0.054	0.004	0.001	0.038	0.038	0.002
$\hat{\beta}_4$	-0.003	0.054	0.055	0.003	-0.001	0.036	0.038	0.002
$\hat{\gamma}_3$	0.004	0.102	0.101	0.005	-0.004	0.73	0.071	0.003
$\hat{\gamma}_4$	-0.003	0.095	0.095	0.005	-0.001	0.067	0.067	0.003
$\hat{\delta}$	-0.008	0.163	0.161	0.018	-0.011	0.112	0.112	0.009

Table 1: *Bias, standard deviation (s.d.), average and interquartile range of the standard errors (ave. s.e., IQR s.e.) for the maximum likelihood estimator of the model parameters. The results are based on 1000 simulated samples with $r = 2$, $T = 4, 8$, $n = 500, 1000$ and $k = 2$.*

are reported in Tables 4 and 5. In particular, Table 4 shows that the bias of the estimator of each parameter is very small, often smaller than the one obtained under the same setting with $r = 2$. As expected, the standard deviation of each estimator slightly increases from $k = 2$ to $k = 3$, but it is always well estimated with the proposed method to compute standard errors. For what concerns the performance of the selection criteria, it may be observed that AIC tends to choose the right number of latent states still with a satisfactory, but consistently lower, probability. On the contrary, BIC performs much better and, in all cases, it led to the correct choice of the number of states with very high frequency.

	$T=4,\ n=500$				$T=4,\ n=1000$			
Est.	Bias	s.d.	ave. s.e.	IQR s.e.	Bias	s.d.	ave. s.e.	IQR s.e.
$\hat{\alpha}_1$	0.003	-	-	-	0.002	-	-	-
$\hat{\beta}_1$	-0.010	0.104	0.105	0.025	0.003	0.078	0.078	0.010
$\hat{\beta}_2$	-0.025	0.100	0.101	0.027	-0.004	0.077	0.078	0.011
$\hat{\gamma}_1$	0.011	0.180	0.178	0.026	0.003	0.140	0.138	0.013
$\hat{\gamma}_2$	-0.038	0.208	0.210	0.024	-0.006	0.144	0.146	0.014
$\hat{\alpha}_2$	0.031	-	-	-	0.007	-	-	-
$\hat{\beta}_3$	0.033	0.114	0.113	0.029	0.012	0.080	0.078	0.011
$\hat{\beta}_4$	-0.036	0.115	0.114	0.026	-0.018	0.081	0.078	0.011
$\hat{\gamma}_3$	-0.034	0.195	0.194	0.032	-0.010	0.135	0.135	0.013
$\hat{\gamma}_4$	0.014	0.195	0.195	0.028	0.003	0.148	0.148	0.013
$\hat{\delta}$	-0.249	0.479	0.480	0.178	-0.050	0.333	0.333	0.051
	$T=8,\ n=500$				$T=8,\ n=1000$			
Est.	Bias	s.d.	ave. s.e.	IQR s.e.	Bias	s.d.	ave. s.e.	IQR s.e.
$\hat{\alpha}_1$	-0.005	-	-	-	-0.001	-	-	-
$\hat{\beta}_1$	0.003	0.069	0.069	0.008	0.002	0.051	0.050	0.004
$\hat{\beta}_2$	-0.003	0.071	0.072	0.008	-0.001	0.054	0.050	0.004
$\hat{\gamma}_1$	-0.008	0.139	0.137	0.010	-0.005	0.093	0.091	0.005
$\hat{\gamma}_2$	0.001	0.125	0.127	0.012	0.004	0.099	0.097	0.006
$\hat{\alpha}_2$	-0.003	-	-	-	-0.001	-	-	-
$\hat{\beta}_3$	-0.008	0.071	0.072	0.008	-0.006	0.050	0.050	0.004
$\hat{\beta}_4$	0.006	0.071	0.073	0.008	0.004	0.052	0.050	0.004
$\hat{\gamma}_3$	-0.002	0.130	0.129	0.011	0.001	0.088	0.090	0.006
$\hat{\gamma}_4$	-0.004	0.128	0.128	0.011	-0.002	0.089	0.090	0.006
$\hat{\delta}$	-0.041	0.265	0.262	0.047	-0.009	0.200	0.199	0.023

Table 2: *Bias, standard deviation (s.d.), average and interquartile range of the standard errors (ave. s.e., IQR s.e.) for the maximum likelihood estimator of the model parameters. The results are based on 1000 simulated samples with* $r = 2$, $T = 4, 8$, $n = 500, 1000$ *and* $k = 3$.

Table 6 (for $r = 2$) and Table 7 (for $r = 3$) show the simulation results concerning the maximum likelihood estimator when samples are generated from the model in which the subject-specific parameters follow a continuous latent process with $n = 500$ and $T = 8$. Under this setting, the number of states k is undefined and then we computed the maximum likelihood estimator of the parameters adopting the value of k chosen with AIC and BIC. The distribution of the predicted k with these two criteria is shown in Table 8.

It may be observed that both AIC and BIC-based estimators perform well, with the former performing better in terms of bias. As may be deduced on the basis of the

T	n	k	Predicted k (AIC) 1	2	3	≥ 4	Predicted k (BIC) 1	2	3	≥ 4
4	500	1	**0.900**	0.091	0.009	0.000	**1.000**	0.000	0.000	0.000
		2	0.000	**0.914**	0.082	0.004	0.009	**0.990**	0.001	0.000
		3	0.000	0.000	**0.898**	0.101	0.000	0.851	**0.149**	0.000
4	1000	1	**0.969**	0.026	0.005	0.000	**1.000**	0.000	0.000	0.000
		2	0.000	**0.941**	0.056	0.003	0.000	**1.000**	0.000	0.000
		3	0.000	0.000	**0.914**	0.086	0.000	0.213	**0.787**	0.000
8	500	1	**0.931**	0.066	0.003	0.000	**1.000**	0.000	0.000	0.000
		2	0.000	**0.918**	0.076	0.006	0.000	**1.000**	0.000	0.000
		3	0.000	0.000	**0.901**	0.099	0.000	0.015	**0.985**	0.000
8	1000	1	**0.988**	0.012	0.000	0.000	**1.000**	0.000	0.000	0.000
		2	0.000	**0.947**	0.052	0.001	0.000	**1.000**	0.000	0.000
		3	0.000	0.000	**0.958**	0.042	0.000	0.000	**1.000**	0.000

Table 3: *Predicted number of latent states with AIC and BIC for the models for $r = 2$ response variables.*

Est.	$k=2$ Bias	s.d.	ave. s.e.	IQR s.e.	$k=3$ Bias	s.d.	ave. s.e.	IQR s.e.
$\hat{\alpha}_1$	0.003	-	-	-	0.009	-	-	-
$\hat{\beta}_1$	0.003	0.052	0.052	0.002	0.003	0.063	0.063	0.004
$\hat{\beta}_2$	-0.003	0.053	0.052	0.002	-0.006	0.063	0.063	0.004
$\hat{\gamma}_1$	0.001	0.095	0.097	0.003	0.001	0.102	0.101	0.005
$\hat{\alpha}_2$	0.001	-	-	-	0.019	-	-	-
$\hat{\alpha}_3$	-0.002	-	-	-	0.002	-	-	-
$\hat{\beta}_3$	0.001	0.043	0.043	0.001	0.004	0.052	0.052	0.003
$\hat{\beta}_4$	-0.001	0.410	0.043	0.001	-0.004	0.053	0.052	0.003
$\hat{\gamma}_2$	0.001	0.081	0.081	0.003	0.001	0.089	0.088	0.004
$\hat{\alpha}_4$	0.009	-	-	-	0.009	-	-	-
$\hat{\alpha}_5$	0.001	-	-	-	-0.012	-	-	-
$\hat{\beta}_5$	-0.001	0.039	0.040	0.001	-0.004	0.051	0.050	0.003
$\hat{\beta}_6$	-0.003	0.042	0.041	0.001	0.004	0.050	0.050	0.003
$\hat{\gamma}_3$	-0.005	0.089	0.088	0.004	-0.008	0.106	0.106	0.006
$\hat{\delta}_1$	0.002	0.114	0.114	0.004	-0.001	0.141	0.142	0.008
$\hat{\delta}_2$	-0.001	0.103	0.105	0.004	-0.002	0.136	0.139	0.008
$\hat{\delta}_3$	0.001	0.121	0.121	0.007	-0.005	0.168	0.170	0.013
$\hat{\delta}_4$	-0.002	0.159	0.160	0.006	-0.011	0.216	0.216	0.017
$\hat{\delta}_5$	-0.006	0.129	0.129	0.004	-0.029	0.280	0.280	0.013
$\hat{\delta}_6$	-0.011	0.238	0.240	0.009	-0.006	0.235	0.235	0.018
$\hat{\delta}_7$	-0.007	0.165	0.161	0.002	-0.014	0.178	0.178	0.013
$\hat{\delta}_8$	0.001	0.152	0.150	0.007	-0.007	0.409	0.407	0.014

Table 4: *Bias, standard deviation (s.d.), average and interquartile range of the standard errors (ave. s.e., IQR s.e.) for the maximum likelihood estimator of the model parameters. The results are based on 1000 simulated samples with $r = 3$, $T = 8$, $n = 500$ and $k = 2, 3$.*

			Predicted k (AIC)				Predicted k (BIC)			
T	n	k	1	2	3	$\geq$4	1	2	3	$\geq$4
4	500	1	**0.880**	0.116	0.004	0.000	**1.000**	0.000	0.000	0.000
		2	0.000	**0.860**	0.131	0.009	0.000	**1.000**	0.000	0.000
		3	0.000	0.000	**0.836**	0.164	0.000	0.079	**0.921**	0.000
4	1000	1	**0.902**	0.098	0.000	0.000	**1.000**	0.000	0.000	0.000
		2	0.000	**0.932**	0.060	0.008	0.002	**0.970**	0.028	0.000
		3	0.000	0.032	**0.902**	0.066	0.000	0.032	**0.968**	0.000
8	500	1	**0.888**	0.090	0.002	0.000	**1.000**	0.000	0.000	0.000
		2	0.000	**0.902**	0.091	0.007	0.000	**1.000**	0.000	0.000
		3	0.000	0.000	**0.858**	0.142	0.000	0.000	**1.000**	0.000
8	1000	1	**0.953**	0.047	0.000	0.000	**1.000**	0.000	0.000	0.000
		2	0.000	**0.926**	0.074	0.000	0.000	**1.000**	0.000	0.000
		3	0.000	0.010	**0.950**	0.040	0.000	0.000	**1.000**	0.000

Table 5: *Predicted number of latent states with AIC and BIC for the models for $r = 3$ response variables.*

	AIC				*BIC*			
Est.	Bias	s.d.	ave. s.e.	IQR s.e.	Bias	s.d.	ave. s.e.	IQR s.e.
$\hat{\alpha}_1$	-0.052	-	-	-	-0.064	-	-	-
$\hat{\beta}_1$	-0.019	0.059	0.055	0.004	-0.032	0.058	0.054	0.004
$\hat{\beta}_2$	0.015	0.056	0.055	0.004	0.032	0.055	0.054	0.004
$\hat{\gamma}_1$	0.027	0.101	0.102	0.006	0.039	0.100	0.100	0.007
$\hat{\gamma}_2$	0.072	0.110	0.108	0.007	0.097	0.112	0.106	0.008
$\hat{\alpha}_2$	-0.054	-	-	-	-0.068	-	-	-
$\hat{\beta}_3$	-0.016	0.058	0.055	0.004	-0.024	0.058	0.054	0.004
$\hat{\beta}_4$	0.016	0.061	0.055	0.004	0.019	0.057	0.054	0.004
$\hat{\gamma}_3$	0.066	0.111	0.109	0.007	0.092	0.111	0.106	0.007
$\hat{\gamma}_4$	0.034	0.107	0.102	0.006	0.041	0.106	0.100	0.006
$\hat{\delta}$	0.166	0.181	0.170	0.026	0.212	0.204	0.161	0.025

Table 6: *Bias, standard deviation (s.d.), average and interquartile range of the standard errors (ave. s.e., IQR s.e.) for the maximum likelihood estimator of the model parameters. The results are based on 1000 simulated samples with $r = 2$, $T = 8$, $n = 500$ and each element of $\boldsymbol{\alpha}_{it}$ following an AR(1). The number of latent states is chosen either with AIC or BIC criterion.*

results in Table 8, this difference is due to the fact that AIC tends to choose a larger number of states than BIC and, with a larger number of states, the continuous latent process is better approximated. Nevertheless, the number of latent states selected with this criterion is small in most cases. The results obtained with other values of n and T are similar to those here shown and confirm that our model can adequately approximate a model based on a continuous latent process of type AR(1), and then

	AIC				BIC			
Est.	Bias	s.d.	ave. s.e.	IQR s.e.	Bias	s.d.	ave. s.e.	IQR s.e.
$\hat{\alpha}_1$	-0.061	-	-	-	-0.191	-	-	-
$\hat{\beta}_1$	0.001	0.056	0.056	0.003	0.001	0.056	0.055	0.003
$\hat{\beta}_2$	-0.009	0.055	0.056	0.003	-0.009	0.055	0.054	0.003
$\hat{\gamma}_1$	0.071	0.113	0.107	0.007	0.071	0.112	0.101	0.004
$\hat{\alpha}_2$	0.023	-	-	-	-0.193	-	-	-
$\hat{\alpha}_3$	-0.063	-	-	-	-0.144	-	-	-
$\hat{\beta}_3$	-0.004	0.048	0.047	0.003	-0.004	0.048	0.045	0.002
$\hat{\beta}_4$	-0.005	0.050	0.046	0.003	-0.005	0.052	0.045	0.002
$\hat{\gamma}_2$	0.051	0.099	0.092	0.006	0.052	0.101	0.089	0.004
$\hat{\alpha}_4$	-0.075	-	-	-	-0.253	-	-	-
$\hat{\alpha}_5$	-0.038	-	-	-	-0.113	-	-	-
$\hat{\beta}_5$	0.003	0.048	0.045	0.003	0.003	0.048	0.049	0.002
$\hat{\beta}_6$	-0.008	0.049	0.045	0.003	-0.008	0.042	0.042	0.002
$\hat{\gamma}_3$	0.072	0.118	0.108	0.012	0.072	0.118	0.099	0.006
$\hat{\delta}_1$	0.047	0.130	0.128	0.009	0.146	0.132	0.128	0.006
$\hat{\delta}_2$	0.053	0.121	0.119	0.007	0.144	0.116	0.119	0.004
$\hat{\delta}_3$	0.056	0.142	0.137	0.013	0.175	0.130	0.137	0.009
$\hat{\delta}_4$	0.049	0.183	0.178	0.016	0.171	0.179	0.179	0.018
$\hat{\delta}_5$	0.035	0.189	0.179	0.029	0.216	0.170	0.179	0.025
$\hat{\delta}_6$	0.060	0.237	0.229	0.044	0.278	0.207	0.229	0.038
$\hat{\delta}_7$	0.051	0.176	0.164	0.027	0.260	0.157	0.164	0.033
$\hat{\delta}_8$	0.073	0.204	0.219	0.040	0.260	0.209	0.219	0.036

Table 7: *Bias, standard deviation (s.d.), average and interquartile range of the standard errors (ave. s.e., IQR s.e.) for the maximum likelihood estimator of the model parameters. The results are based on 1000 simulated samples with* $r = 3$, $T = 8$, $n = 500$ *and each element of* $\boldsymbol{\alpha}_{it}$ *following an* $AR(1)$. *The number of latent states is chosen either with AIC or BIC criterion.*

	Predicted k (AIC)					Predicted k (BIC)				
r	1	2	3	4	≥ 5	1	2	3	4	≥ 5
2	0.000	0.018	0.885	0.088	0.009	0.000	0.122	0.878	0.000	0.000
3	0.000	0.000	0.027	0.804	0.170	0.000	0.000	0.955	0.045	0.000

Table 8: *Predicted number of latent states for the case of* $r = 2$ *response variables considered in Table 6 and for that of* $r = 3$ *response variables considered in Table 7.*

reliable parameter estimates can be obtained. Obviously, we need to be cautious in generalizing this conclusion to continuous latent processes of a different nature. For instance, we expect that the approximation can be inadequate in the presence of an AR(2) process, which has a dependence structure different from the one assumed in our model.

6 Analysis of the PSID dataset

We illustrate the proposed model through the analysis of a dataset which is very similar to that used in the study of Hyslop (1999). The dataset was extracted from the database deriving from the Pseudo Study of Income Dynamics, which is primarily sponsored by the National Science Foundation, the National Institute of Aging, and the National Institute of Child Health and Human Development and is conducted by the University of Michigan. This database is freely accessible from the website `http://psidonline.isr.umich.edu`, to which we refer for details.

Our dataset concerns $n = 1446$ women who were followed from 1987 to 1993. There are two binary response variables: *fertility* (indicating whether a woman had given birth to a child in a certain year) and *employment* (indicating whether she was employed). The covariates are: *race* (dummy variable equal to 1 for a black woman), *age* (in 1986), *education* (year of schooling), *child 1-2* (number of children in the family aged between 1 and 2 years, referred to the previous year), *child 3-5*, *child 6-13*, *child 14-*, *income of the husband* (in dollars, referred to the previous year).

In analyzing the dataset, the most interesting scientific question concerns the direct effect of fertility on employment. Also of interest are the strength of the state dependence effect for both response variables and how these variables depend on the covariates. The proposed approach allows us to separate these effects from the effect of the unobserved heterogeneity by modeling the latter by a latent process. In this way, we admit that the unobserved heterogeneity effect on the response variables is time-varying; this is not allowed either within a latent class model with covariates or in the most common random effect models.

On these data, we fitted the proposed model with a number of latent states k from 1 to 5. The model is formulated on the basis of assumptions (4) and (5), with $\boldsymbol{X}_{it} = \boldsymbol{1}_2 \otimes \boldsymbol{x}'_{it}$ and $\boldsymbol{Y}_{it} = \boldsymbol{1}_2 \otimes \boldsymbol{y}'_{i,t-1}$, $t = 1, \ldots, T$, and on assumption (9), with $\boldsymbol{Y}_{i0} = \boldsymbol{1}_{k-1} \otimes \begin{pmatrix} 1 & \boldsymbol{y}'_{i0} \end{pmatrix}$. The vector $\boldsymbol{x}_{it}$ includes the covariates indicated above further to a dummy variable for each year. The results of this preliminary analysis are reported in Table 9 in terms of maximum log-likelihood, AIC and BIC. For each value of k we adopted both the deterministic and the random search mechanism described at the end of Section 4.1 to initialize the EM algorithm and we report the results corresponding to the best solution in terms of likelihood, provided that the corresponding observed information matrix $\boldsymbol{J}(\hat{\boldsymbol{\theta}})$ is of full rank. In Table 9 we also report the computing time needed to run, on a Sun XFire 4100 computer with AMD dual-core Opteron and 8GB RAM, our MATLAB implementation of the EM algorithm (with the deterministic starting rule) and of the procedure for computing the standard errors. This computing time is reasonable considering the complexity of the dataset and the fact that we do not adopt an optimized programming code. Further, since three is the proper number of latent states for these data, the computing time considerably increases when fitting a model with a larger number of states. We note, instead, that there is not much increase in computing time when passing from four to five latent states.

	k				
	1	2	3	4	5
log-lik.	-6219.0	-6050.0	-6011.5	-6004.7	-5993.6
# par.	37	44	53	64	77
AIC	12512	12188	12129	12137	12141
BIC	12707	12420	12409	12475	12548
Time	37s	3m21s	15m59s	1h19m41s	1h40m21s

Table 9: *Log-likelihood, number of parameters, AIC, BIC and computing time resulting from fitting the proposed latent Markov model with 1 to 5 latent states.*

On the basis of these results, we conclude that $k = 3$ is a suitable number of latent states for the PSID dataset; in fact, this value of k corresponds to the minimum value of both AIC and BIC indices.

In Table 10 we show the estimates of the parameters affecting the marginal logits of fertility and employment and the log-odds ratio between these variables, again for k from 1 to 5. We recall that these parameters are collected in vectors $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$.

		k				
	Effect	1	2	3	4	5
logit fertility	intercept*	-1.807	-2.072	**-2.117**	-2.198	-2.101
	race	-0.230**	-0.230**	**-0.235****	-0.243**	-0.239**
	age†	-0.216**	-0.218**	**-0.223****	-0.226**	-0.224**
	$(\text{age}^\dagger)^2/100$	-1.112**	-1.122**	**-1.135****	-1.153**	-1.107**
	education†	0.152**	0.154**	**0.160****	0.162**	0.160**
	child 1-2	0.183**	0.187**	**0.177****	0.177**	0.170**
	child 3-5	-0.360**	-0.374**	**-0.389****	-0.390**	-0.388**
	child 6-13	-0.594**	-0.605**	**-0.611****	-0.613**	-0.608**
	child 14-	-0.879**	-0.885**	**-0.893****	-0.897**	-0.903**
	income†/1000	0.002	0.002	**0.002**	0.002	0.002
	lag fertility	-1.476**	-1.469**	**-1.482****	-1.452**	-1.499**
	lag employment	-0.163	0.212	**0.444****	0.443**	0.427**
logit employment	intercept*	-0.688	0.523	**-0.010**	-0.205	0.087
	race	0.099	0.125	**0.134**	0.163	0.192
	age†	0.015**	0.028	**0.068****	0.070**	0.074**
	$(\text{age}^\dagger)^2/100$	-0.103	-0.093	**0.045**	0.109	-0.205
	education†	0.102**	0.125	**0.096****	0.104**	0.121**
	child 1-2	-0.116**	-0.174	**-0.089**	-0.010	-0.031
	child 3-5	-0.234**	-0.219	**-0.190****	-0.1613	-0.146
	child 6-13	-0.062	0.012	**-0.006**	0.030	0.034
	child 14-	-0.010	0.052	**0.065**	0.086	0.160
	income†/1000	-0.009**	-0.009	**-0.013****	-0.013**	-0.014**
	lag fertility	-0.478**	-0.733**	**-0.704****	-0.654**	-0.747**
	lag employment	2.949**	1.571**	**1.008****	1.079**	0.746**
log-odds ratio	intercept	-1.213**	-1.286**	**-1.130****	-1.651**	-1.173**

Table 10: *Maximum likelihood estimates of the model parameters affecting the marginal logits for fertility and employment and the log-odds ratio (*average of the support points based on the posterior probabilities, †minus the sample average, **significant at the 5% level, in boldface the parameter estimates for the selected model).*

On the basis of the estimates of the parameters for the covariates under the selected number of states $k = 3$, we conclude that race has significant effect on fertility. In fact, as shown in Table 10, the estimate of the coefficient for the corresponding dummy is equal to -0.235 with a p-value less than 0.05. On the other hand, this covariate has not a significant effect on employment. Similarly, age seems to have a stronger effect on fertility than on employment. At this regard consider that the women in the sample were aged between 18 and 47, which is a limited range of years if we want to effectively study the effect of aging on the probability of having a job position. Other considerations arising from Table 10 are that education has a significant effect on both fertility and employment, whereas the number of children in the family strongly affects only the first response variable and income of the husband strongly affects only the second one. Very interesting are the estimates of the association parameters, i.e. the log-odds ratio between the two response variables and the parameters measuring the effect of the lagged responses on the marginal logits. The log-odds ratio is negative and highly significant, meaning that the response variables are negatively associated when referred to the same year. On the other hand, lagged fertility has a significant negative effect on both response variables, whereas lagged employment has a significant negative effect on the first variable and a significant positive effect on the second variable. These estimates allow us to conclude that fertility has a negative effect on the probability of having a job position in the same year of the birth and the following one, whereas employment is serially positively correlated (as consequence of the state dependence effect) and fertility is negatively serially correlated.

For the model based on $k = 3$ latent states, we also show in Table 11 the estimates of the support points (one for the marginal logit of fertility and the other for that of employment) corresponding to each latent state, the estimates of the parameters $\boldsymbol{\phi}$ of the model on the initial probabilities of the latent states, and the estimated transition probability matrix. We recall that we assume a multinomial logit model on these probabilities, with the first latent state taken as reference category. This model uses, as covariates, fertility and employment at the initial year of observation; see assumption (9). The corresponding initial probabilities of the three states, averaged on all the subjects in the samples, are equal to 0.100, 0.266 and 0.634, respectively. Further, the average probability of each latent state at every time occasion is represented in Figure 1.

Latent state	Support points		Initial prob. parameters			Transition probabilities		
	Fertility	Empl.	Intercept	Fertility	Empl.			
1	-1.349	-5.358	-	-	-	0.947	0.050	0.003
2	-1.858	-1.066	0.775	0.337	0.861	0.068	0.888	0.044
3	-2.505	2.205	0.370	0.015	4.253**	0.003	0.092	0.906

Table 11: *Estimated support points for each latent state, estimated parameters for the corresponding initial probabilities and estimated transition probability matrix.*

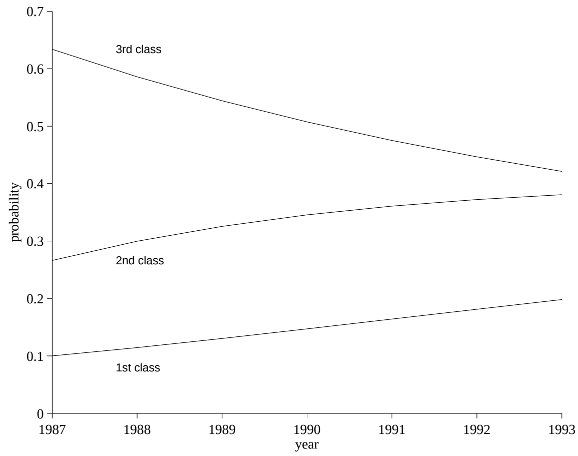

Figure 1: *Estimated average probability of each latent state at every time occasion.*

As may be deduced looking at the estimates for the support points in Table 11, the three latent states correspond to different levels of propensity to give birth to a child and to have a job position. The first latent state, with support point $\hat{\boldsymbol{\xi}}_1 = (-1.349, -5.358)'$, corresponds to subjects with the highest propensity to fertility and the lowest propensity to have a job position. In fact, the first element of $\hat{\boldsymbol{\xi}}_1$ is higher and the second is lower than the corresponding elements of the other support points $\hat{\boldsymbol{\xi}}_2$ and $\hat{\boldsymbol{\xi}}_3$. On the contrary, the third latent state corresponds to subjects with the lowest propensity to fertility and the highest propensity to have a job position. Finally, the second state is associated to intermediate levels of both propensities. It is also interesting to observe that the transition matrix has an almost symmetric structure which implies the evolution of the probability of each state represented in Figure 1. We can note that the probability of the first two latent states grows across time, whereas that of the third latent state decreases, but this state always remains the one with highest probability. The consequence is that women without children and not having a job position in the previous year tend to become more inclined to childbearing and less inclined to have a job position as time goes.

In order to better investigate the features of the latent process, we also tested the hypothesis that the transition matrix is diagonal, so that a latent class model with covariates results. The latter may be fitted by a simpler version of the EM algorithm illustrated in Section 4.1. The likelihood ratio statistic for this hypothesis is equal to 40.848 which, on the basis of the results of Bartolucci (2006), leads us to strongly reject the hypothesis. In order to help the comparison between the proposed model and its latent class version, we also report in Tables 12 and 13 a summary of the results obtained with the latter, for a number of latent classes k between 1 and 5.

	k				
	1	2	3	4	5
log-lik.	-6219.0	-6064.3	-6031.7	-6025.1	-6022.7
# par.	37	42	47	52	57
AIC	12512	12213	12157	12154	12159
BIC	12707	12434	12405	12429	12460

Table 12: *Log-likelihood, number of parameters, AIC and BIC resulting from fitting the latent class version of the proposed model with 1 to 5 latent classes.*

It is worth noting that the smallest value of the AIC index obtained with the proposed model is smaller than that reachable with its latent class version. This confirms that, realistically, the effect of unobservable characteristics of a subject on fertility and employment is not time-constant. The implications of ignoring this aspect may be deduced by comparing the parameter estimates in Table 13 with those in Table 10. The most evident difference is in the effect of lagged employment on the marginal

		k				
	Effect	1	2	3	4	5
logit fertility	intercept*	-1.807	-1.900	-1.988	-1.921	-2.881
	race	-0.230**	-0.226	-0.241**	-0.245**	-0.248**
	age†	-0.216**	-0.216**	-0.217**	-0.218**	-0.222**
	(age†)^2/100	-1.112**	-1.126**	-1.127**	-1.147**	-1.167**
	education†	0.152**	0.152**	0.153**	0.151**	0.155**
	child 1-2	0.183**	0.187**	0.183**	0.156**	0.080
	child 3-5	-0.361**	-0.369**	-0.379**	-0.390**	-0.428**
	child 6-13	-0.594**	-0.603**	-0.613**	-0.616**	-0.638**
	child 14-	-0.879**	-0.883**	-0.889**	-0.893**	-0.909**
	income†/1000	0.002	0.002	0.002	0.002	0.003
	lag fertility	-1.476**	-1.459**	-1.462**	-1.503**	-1.575**
	lag employment	-0.163	-0.018	0.118	0.034	0.005
logit employment	intercept*	-0.688	0.014	-0.143	-1.043	-0.630
	race	0.099	0.082	0.160	0.181	0.180
	age†	0.015**	0.016	0.021	0.021	0.021
	(age†)^2/100	-0.103	0.010	0.002	-0.011	-0.014
	education†	0.102**	0.119**	0.116**	0.124**	0.126**
	child 1-2	-0.116**	-0.177**	-0.123	-0.182**	-0.178**
	child 3-5	-0.234**	-0.170**	-0.159**	-0.190**	-0.186**
	child 6-13	-0.062	0.046	0.051	0.058	0.062
	child 14-	-0.010	0.048	0.050	0.064	0.068
	income†/1000	-0.009**	-0.009**	-0.010**	-0.010**	-0.010**
	lag fertility	-0.478**	-0.681**	-0.617**	-0.677**	-0.680**
	lag employment	2.949**	2.061**	1.791**	1.751**	1.753**
log-odds ratio	intercept	-1.213**	-1.302**	-1.227**	-1.300**	-1.325**

Table 13: *Estimates of the parameters affecting the marginal logits for fertility and employment and the log-odds ratio under the latent class version of the proposed model (*average of the support points based on the posterior probabilities, †minus the sample average, **significant at the 5% level).*

logit of this response variable. The estimate of this effect never goes below 1.751 under the latent class model, which is much higher than the value obtained under the proposed model, corresponding to 1.008. Then, a model which ignores that the effect of unobserved heterogeneity might be time-varying usually leads to an overestimation of the state dependence effect with, for example, important consequences on the evaluation of the opportunity of an employment policy.

Finally, for each woman in the sample we estimated the a posteriori most likely sequence of latent states by using the Viterbi algorithm. As an illustration, consider a white woman in the sample who was 27 years old in 1986, with 12 years of education and no children in the same year, and having a husband with income between 10,000 and 21,000 dollars in the period of interest. This woman had no children in 1987 and 1993, and had a job position in 1987 and 1988 and continuously from 1991 to 1993. The corresponding predicted sequence of latent states is 3, 3, 2, 2, 2, 2, 2, meaning that this woman was in the third state in 1987 and 1988 and then she moved to the second. Consequently, her propensity to childbearing has increased across time.

Overall, it results that 78.5% of the women started and persisted in the same latent state for the entire period, whereas for the 21.5% of the women we had one or more transitions between states. The presence of these transitions explains the difference between the estimates of the association parameters under the proposed latent Markov model (see Table 10) and its latent class version (see Table 13).

7 Discussion

In this paper, we extend the dynamic logit model (Hsiao, 2005) for binary longitudinal data in two directions. First, we allow modeling response variable vectors with any number and any kind of categorical responses. Second, we allow for the presence of subject-specific parameters which are time-varying and follow a first-order Markov chain which is not directly observable. The resulting model may be considered as a transition model (Molenberghs and Verbeke, 2004) for multivariate categorical longitudinal data, since the responses at a certain occasion are also modeled conditional on their values at the previous occasion. The approach is then different from approaches in which the marginal distribution of the response variables at each occasion is directly modeled; see, for instance, Lang and Agresti (1994) and Molenberghs and Lesaffre (1994). However, at least in our context of application, we consider transition models more interesting since they allow one to directly measure the state dependence effect (Heckman, 1981b), i.e. the real effect that experiencing a certain situation in the present has on the probability of experiencing the same situation in the future.

Two features of the proposed approach are worth to be remarked. First, the approach relies on a flexible family of link functions to parameterize in a meaningful way the conditional distribution of the vector of response variables. This family is based on marginal logits and log-odds that may be of different types so as to suit at best the nature of the data. For instance, global or continuation logits and log-odds ratios may be used with ordinal response variables. Second, by assuming that the latent process is discrete we avoid parametric assumptions on it, giving in this way more flexibility to the resulting model in the sense of Heckman and Singer (1984) and Lindsay et al. (1991). Assuming a discrete instead of a continuous latent process also has the advantage of permitting to exactly compute the likelihood of the model without requiring quadrature or Monte Carlo methods. On the other hand, some simulation results illustrated in Section 5 show that the maximum likelihood estimator of the parameters has a reduced bias even when data are generated from a version of the model based on a continuous latent process. However, these results have to be cautiously taken considering that they come from a rather limited simulation study in which the true model is based on an AR(1) process. A drawback of assuming a discrete latent process is that the number of model parameters quickly increases with the number of latent states. Though these simulation results confirm that a small number of states is often required in order to have an adequate fit, the model may be made more parsimonious by imposing suitable constraints on the transition matrix.

Another aspect to be remarked concerns the numerical complexity of the EM algorithm for computing the maximum likelihood estimate of the model parameters. As for standard latent variable models, this algorithm may require a large number of steps. However, in the simulation study and in our application we did not observe particular problems of instability or lack of convergence. Moreover, as the number of response variables or its categories increases, the numerical complexity of the algorithm grows at a reasonable rate. This is because we rely on a parametrization of the distribution of the response variables based on effects (marginal logits and log-odds ratios) whose number does not increase exponentially with the number of these variables. Moreover, the EM algorithm did not show particular problems with either a large number of states or a large number of time occasions. This is because we use special recursions to exactly compute the likelihood and the conditional probabilities of the latent states required within this algorithm. Moreover, we observed that the number of iterations required to reach the convergence of the EM algorithm tends to be small when data are generated from a model based on a limited number of well separated latent states. On the other hand, special care has to be payed in order to check that the point at convergence of the algorithm corresponds to the global maximum of the likelihood. For this aim, we suggested a procedure based on a deterministic and a random rule for choosing the starting values for this algorithm which seems to work properly.

A final point concerns possible extensions of the proposed approach. A simple extension consists in allowing the number of time occasions to vary between subjects. Though not explicitly showed, this extension may be simply implemented in our approach by adapting to this case the recursions illustrated in Appendix. The structure of the EM algorithm illustrated in Section 4.1 does not need any relevant adjustment. Though some adjustments to the estimation algorithm are necessary, the model may also be used when a different number of response variables is observed between occasions. This is made possible by the adopted parametrization which gives rise to the same interpretation for the parameters of interest regardless of the number of response variables. In fact, it is based on marginal effects which, when referred to the same set of response variables, are always expressed in the same way. This feature is not shared by parametrizations of log-linear type, which are based on conditional logits and higher order interactions given a reference value of the other variables.

Appendix: marginal and posterior probabilities

Efficient computation of the probability in (10) may be performed by exploiting a forward recursion available in the hidden Markov literature, and which is here expressed by using the matrix notation; see also MacDonald and Zucchini (1997) and Bartolucci (2006).

The recursion consists of computing, for $t = 1, \dots, T$, the vector

$$\boldsymbol{q}_{it}(\boldsymbol{y}_{i1}, \dots, \boldsymbol{y}_{it}) = \begin{cases} \mathrm{diag}[\boldsymbol{u}_{i1}(\boldsymbol{y}_{i1})]\boldsymbol{\lambda}(\boldsymbol{y}_{i0}) & \text{if } t = 1, \\ \mathrm{diag}[\boldsymbol{u}_{it}(\boldsymbol{y}_{it})]\boldsymbol{\Pi}'\boldsymbol{q}_{it}(\boldsymbol{y}_{i1}, \dots, \boldsymbol{y}_{i,t-1}) & \text{otherwise,} \end{cases}$$

where $\boldsymbol{u}_{it}(\boldsymbol{y}_{it})$ is a column vector with elements $p(\boldsymbol{y}_{it}|\boldsymbol{\alpha}_{it} = \boldsymbol{\xi}_c, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})$, $c = 1, \dots, k$. We then compute $p(\boldsymbol{y}_{i1}, \dots, \boldsymbol{y}_{iT}|\boldsymbol{x}_{i1}, \dots, \boldsymbol{x}_{iT}, \boldsymbol{y}_{i0})$ as the sum of the elements of $\boldsymbol{q}_{iT}(\boldsymbol{y}_{i1}, \dots, \boldsymbol{y}_{iT})$.

For what concerns the posterior probabilities in (11) and (12), let $\boldsymbol{V}_{it}(\boldsymbol{y}_{i1}, \dots, \boldsymbol{y}_{iT})$ be a matrix with elements $p(\boldsymbol{\alpha}_{i,t-1} = \boldsymbol{\xi}_c, \boldsymbol{\alpha}_{it} = \boldsymbol{\xi}_d|\boldsymbol{x}_{i1}, \dots, \boldsymbol{x}_{iT}, \boldsymbol{y}_{i0}, \dots, \boldsymbol{y}_{iT})$ for $c, d = 1, \dots, k$. For $t = 2, \dots, T$, this matrix may be computed as follows

$$\boldsymbol{V}_{it}(\boldsymbol{y}_{i1}, \dots, \boldsymbol{y}_{iT}) = \frac{\mathrm{diag}[\boldsymbol{q}_{i,t-1}(\boldsymbol{y}_{i1}, \dots, \boldsymbol{y}_{i,t-1})]\boldsymbol{\Pi}\mathrm{diag}[\boldsymbol{u}_{it}(\boldsymbol{y}_{it})]\mathrm{diag}[\boldsymbol{v}_{it}(\boldsymbol{y}_{it}, \dots, \boldsymbol{y}_{iT})]}{p(\boldsymbol{y}_{i1}, \dots, \boldsymbol{y}_{iT}|\boldsymbol{x}_{i1}, \dots, \boldsymbol{x}_{iT}, \boldsymbol{y}_{i0})},$$

where the vector $\boldsymbol{v}_{it}(\boldsymbol{y}_{it}, \dots, \boldsymbol{y}_{iT})$ is equal to $\boldsymbol{1}_k$ for $t = T$ and, for $t < T$, is computed as $\boldsymbol{\Pi}\mathrm{diag}[\boldsymbol{u}_{i,t+1}(\boldsymbol{y}_{i,t+1})]\boldsymbol{v}_{i,t+1}(\boldsymbol{y}_{i,t+1}, \dots, \boldsymbol{y}_{iT})$. The probabilities $p(\boldsymbol{\alpha}_{it} = \boldsymbol{\xi}_c|\boldsymbol{x}_{i1}, \dots, \boldsymbol{x}_{iT}, \boldsymbol{y}_{i0}, \dots, \boldsymbol{y}_{iT})$ may then be computed by suitable sums of the elements of $\boldsymbol{V}_{it}(\boldsymbol{y}_{i1}, \dots, \boldsymbol{y}_{iT})$.

References

Bartolucci, F. (2006). Likelihood inference for a class of latent markov models under linear hypotheses on the transition probabilities. *Journal of the Royal Statistical Society, Series B*, 68:155–178.

Bartolucci, F., Colombi, R., and Forcina, A. (2007a). An extended class of marginal link functions for modelling contingency tables by equality and inequality constraints. *Statistica Sinica*, 17:691–711.

Bartolucci, F. and Nigro, V. (2007). Maximum likelihood estimation of an extended latent markov model for clustered binary panel data. *Computational Statistics and Data Analysis*, 51:3470–3483.

Bartolucci, F., Pennoni, F., and Francis, B. (2007b). A latent markov model for detecting patterns of criminal activity. *Journal of the Royal Statistical Society, series A*, 170:115–132.

Carrasco, R. (2001). Binary choice with binary endogenous regressors in panel data: estimating the effect of fertility on female labor participation. *Journal of Business and Economic Statistics*, 19:385–394.

Colombi, R. and Forcina, A. (2001). Marginal regression models for the analysis of positive association of ordinal response variables. *Biometrika*, 88:1007–1019.

Dardanoni, V. and Forcina, A. (1998). A unified approach to likelihood inference on stochastic orderings in a nonparametric context. *Journal of the American Statistical Association*, 93:1112–1123.

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

Glonek, G. F. V. (1996). A class of regression models for multivariate categorical responses. *Biometrika*, pages 15–28.

Glonek, G. F. V. and McCullagh, P. (1995). Multivariate logistic models. *Journal of the Royal Statistical Society, Series B*, 57:533–546.

Heckman, J. (1981a). The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. In Manski, C. and McFadden, D., editors, *Structural Analysis of Discrete Data*, Cambridge, MA. MIT press.

Heckman, J. and Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica*, 52:271–320.

Heckman, J. J. (1981b). Heterogeneity and state dependence. In L., M. D. and A, M. C., editors, *Structural Analysis of Discrete Data*, Cambridge, MA. MIT Press.

Hsiao, C. (2005). *Analysis of Panel Data*. Cambridge University Press, New York.

Hyslop, D. R. (1999). State dependence, serial correlation and heterogeneity in intertemporal labor force participation of married women. *Econometrica*, 67:1255–1294.

Juang, B. and Rabiner, L. (1991). Hidden Markov models for speech recognition. *Technometrics*, 33:251–272.

Lang, J. and Agresti, A. (1994). Simultaneously modeling joint and marginal distributions of multivariate categorical responses. *Journal of the American Statistical Association*, 89:625–632.

Lindsay, B., Clogg, C., and Grego, J. (1991). Semiparametric estimation in the Rasch model and related exponential response models, including a simple latent class model for item analysis. *Journal of the American Statistical Association*, 86:96–107.

MacDonald, I. L. and Zucchini, W. (1997). *Hidden Markov and other Models for Discrete-Valued Time Series*. Chapman and Hall, London.

McCullagh, P. (1980). Regression models for ordinal data (with discussion). *Journal fo the Royal Statistical Society, Series B*, 42:109–142.

McLachlan, G. and Peel, D. (2000). *Finite Mixture Models*. Wiley.

Molenberghs, G. and Lesaffre, E. (1994). Marginal modelling of correlated ordinal data using a multivariate Plackett distribution. *Journal of the American Statitical Association*, 89:633–644.

Molenberghs, G. and Verbeke, G. (2004). Meaningful statistical model formulations for repeated measures. *Statistica Sinica*, 14:989–1020.

Shapiro, A. (1988). Towards a unified theory of inequality constrained testing in multivariate analysis. *International Statistical Review*, 56:49–62.

Silvapulle, M. and Sen, P. (2004). *Constrained Statistical Inference*. Wiley.

Ten Have, T. R. and Morabia, A. (1999). Mixed effects models with bivariate and univariate association parameters for longitudinal bivariate binary response data. *Biometrics*, 55:85–93.

Todem, D., Kim, K., and Lesaffre, E. (2007). Latent-variable models for longitudinal data with bivariate ordinal outcomes. *Statistics in Medicine*, 26:1034–1054.

Vermunt, J. K., Langeheine, R., and Böckenholt, U. (1999). Discrete-time discrete-state latent markov models with time-constant and time-varying covariates. *Journal of Educational and Behavioral Statistics*, 24:179–207.

Viterbi, A. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 13:260–269.

Wiggins, L. (1973). *Panel Analysis: Latent probability models for attitude and behavious processes*. Elsevier, Amsterdam.