Bias-adjusted three-step latent Markov modeling with covariates

Roberto Di Mari

Department of Economics and Finance, University of Rome Tor Vergata, Italy

Daniel L. Oberski

Jeroen K. Vermunt

Department of Methodology and Statistics, Tilburg University, The Netherlands

Author Note

Address:

Room P1 S39,

via Columbia, 2, 00113 Rome

Phone: +390672595932

Email: roberto.di.mari@uniroma2.it

Abstract

Latent Markov models with covariates can be estimated via one-step Maximum Likelihood. However, this one-step approach has various disadvantages, such as that the inclusion of covariates in the model may alter the formation of the latent states and that parameter estimation may become infeasible with large numbers of time points, responses, and covariates. This is why researchers typically prefer performing the analysis in a stepwise manner; that is, they first construct the measurement model, then obtain the latent state classifications, and subsequently study the relationship between covariates and latent state memberships. However, such a stepwise approach yields downward biased estimates of the covariate effects on initial state and transition probabilities. The present paper shows how to overcome this problem using a generalization of the bias-corrected three-step estimation method proposed for latent class analysis (Asparouhov & Muthén, 2014; Bolck, Croon, & Hagenaars, 2004; Vermunt, 2010). We give a formal derivation of the generalization to latent Markov models and discuss how it can be used with many time points by incorporating it into a Baum-Welch type of EM algorithm. We evaluate the method through a simulation study and illustrate it using an application on household financial portfolio change. Our study shows that the proposed correction method yields unbiased parameter estimates and accurate standard errors, except for situations with very poorly separated classes and a small sample.

Keywords: latent Markov model, hidden Markov model, three-step approach, bias-correction, latent transition analysis, latent class analysis, classify-analyze

Bias-adjusted three-step latent Markov modeling with covariates

Introduction

The latent Markov (LM) model (Collins & Wugalter, 1992; Van de Pol & De Leeuw, 1986; Van de Pol & Langeheine, 1990; Wiggings, 1973), also referred to as hidden Markov or latent transition model, is a variant of the latent class model (Goodman, 1974a and 1974b; Lazarsfeld, 1950; Lazarsfeld & Henry, 1968; Vermunt & Magidson, 2002) for the analysis of longitudinal data. It allows researchers to analyze how individuals change between latent classes or latent states over time. An important extension of the basic LM model is a model which allows the inclusion of time-constant and time-varying covariates explaining differences in initial states and transitions across individuals and/or time points (Collins & Lanza, 2010; Vermunt, Langeheine, & Böckenholt, 1999).

In practice, however, researchers encounter several obstacles to fitting LC models with covariates (Asparouhov & Muthén, 2014; Vermunt, 2010). The correct latent class (measurement) model is not usually known in advance, so that a large number of LC models needs to be investigated, as with regular latent class analysis. However, with LM covariate models this number increases exponentially through combination with other choices such as the structural lag and the specification of the covariate effects, including functional form and variable selection. This makes model selection with LM models an especially daunting task. To circumvent these issues, a practical approach is sometimes taken in which, first, the measurement model without covariates is selected, and, subsequently, the covariates are included. However, when adding the covariates to the structural model, the measurement part of the model, including class definitions and the optimal number of classes, can change, negating the simplification. Moreover, with many time points, indicators, and covariates, each step of this analysis can take a considerable amount of time. For these reasons, researchers often prefer a “three-step” (“classify-analyze”) approach: after estimating a model without covariates (step 1), latent state assignments are obtained (step 2), and, subsequently, the relationship between the assigned state memberships and covariates is modeled (step 3).

However, as shown by Bolck et al. (2004), the three-step approach yields biased estimates of covariate effects. Several authors have shown how to remove this bias in standard

LC models (Asparouhov & Muthén, 2014; Bakk, Tekle, & Vermunt, 2013; Bolck et al. 2004; Vermunt, 2010). More specifically, Bolck et al. (2004) developed a bias correction method (referred to as the BCH method) which involves performing a logistic regression on the reweighted data set. Vermunt (2010) proposed an alternative maximum likelihood based bias correction (referred to as the ML method) which involves estimating a specific type of LC model in the third step, which incorporates the classification error in the class assignments explicitly.

The aim of the current study is to generalize the ML correction method for three-step estimation to latent Markov models with covariates. The LM model is also an LC model, and it therefore seems intuitive that bias-corrected three step approaches could be applied to LM modeling as well (e.g. Asparouhov & Muthén, 2014). However, in doing so, several analytical and practical challenges remain, which we aim to resolve. First, the BCH and ML methods apply to models with a single latent class variable. With multiple latent variables, to simply apply the existing results would require considering the full joint distribution of all latent variables, which would be infeasible in models with many time points. Fortunately, however, we show that the step three model factors into a single-indicator LM model, considerably simplifying estimation. Second, a practical issue with LM models is that there are essentially as many measurement models as there are time points. With many time points this may make the analysis highly challenging. However, we show that pooling the time points in the first step gives good results, simplifying the problem into a single standard LC model. Third and finally, the empirical performance of bias-corrected three step estimation in LM models has not been investigated thus far. In a series of simulations, we demonstrate that this performance is adequate under moderate assumptions.

Recently, Bartolucci, Montanari, & Pandolfi (2015) proposed another type of stepwise procedure for LM models with covariates. They estimate a LC model (measurement part) on the pooled data in the first step, from which they get posterior membership probabilities (second step). In the third step, they estimate the Markov part of the model using a pairwise likelihood function for transitions between adjacent time point, while fixing the measurement part of the model. Their simulation study showed that this approach works well only when class separation in the first step is close to perfect. We expect that our generalization of ML type correction will also work in situations in which class separation is not very good.

The remainder of the paper is organized as follows. In Section 2 we introduce the LM model with covariates. Section 3 discusses the three-step method, which includes the two alternatives for step one, the correction method used to account for classification error in step two, and the final step of the procedure. In Section 4 we report the results of a simulation study in which we compared the correction methods under different conditions. Section 5 presents an empirical application, and Section 6 concludes with a final discussion and possible directions for future research.

Latent Markov Modeling with Covariates

Let T denote the number of time occasions and let $Y_{1t}, \dots, Y_{Jt}$ be J indicator variables measured at time t, and let us indicate as $\boldsymbol{Y}_t$ a full response pattern at time $t = 0, \dots, T$, and its realization $\boldsymbol{y}_t$; $\boldsymbol{Y}$ denotes the full set of response configurations at all $T+1$ occasions. Let us assume a set of covariates is available; we denote $\boldsymbol{Z}_t$ the vector of K covariates at time t and we indicate the full set of covariate values at all $T+1$ occasions as $\boldsymbol{Z}$. The categorical latent variable at time t is referred to as X_t and we let s_t be one of the S possible latent states for time occasion t. A latent Markov model specifies the probability of having a particular sequence of response configurations at $T+1$ time occasions, given the time-specific covariates, and is formulated as

$$P(\boldsymbol{Y} = \boldsymbol{y}|\boldsymbol{Z}) = \sum_{s_0=1}^{S} \sum_{s_1=1}^{S} \cdots \sum_{s_T=1}^{S} P(X_0 = s_0|\boldsymbol{Z}_0) \prod_{t=1}^{T} P(X_t = s_t|X_{t-1} = s_{t-1}, \boldsymbol{Z}_t) \prod_{t=0}^{T} P(\boldsymbol{Y}_t = \boldsymbol{y}_t|X_t = s_t). \tag{1}$$

The probability $P(\boldsymbol{Y} = \boldsymbol{y}|\boldsymbol{Z})$ is defined by two components. A so-called structural component, which describes changes in latent state across time points, and a measurement component, connecting the latent state at a particular point in time to the observed responses.

The presence of the covariates affects the model only through the structural component, for which we assume a first-order Markov model. That is, when controlling for covariates at time t, X_t is only affected by X_{t-1}. This involves modeling the initial latent state probability $P(\boldsymbol{X}_0 = s_0|\boldsymbol{Z}_0)$, and the latent transition probability $P(\boldsymbol{X}_t = s_t|\boldsymbol{X}_{t-1} = s_{t-1}, \boldsymbol{Z}_t)$, respectively given the covariate values at the initial state and at time t.

The measurement part has the form of a standard latent class model for dichotomous response variables (Goodman, 1974a and 1974b; Hagenaars, 1990). Under the assumption of local independence of response variables given class membership, the conditional distribution of the responses can be written as

$$P(\boldsymbol{Y}_t = \boldsymbol{y}_t | X_t = s_t) = \prod_{j=1}^{J} P(Y_{jt} = y_{jt} | X_t = s_t). \tag{2}$$

Figure 1 depicts the LM model graph as in Equation (1). The present paper focuses on binary response variables. In this case it is natural to assume that, given the latent variable, Y_{jt} has a Bernoulli distribution with a certain success probability π^t_{sj}. In addition, in the general formulation of the model, the conditional distribution of $\boldsymbol{Y}_t$ is allowed to differ across time.

For each combination of item and time, $1 \leq j \leq J$ and $0 \leq t \leq T$, the conditional response probabilities can be parametrized through log-odds in the following manner

$$\log\left(\frac{\pi^t_{sj}}{1-\pi^t_{sj}}\right) = \theta^t_{js}, \quad \text{for } 1 \leq s \leq S. \tag{3}$$

Here we consider a time-invariant measurement model. In addition, one might wish to include covariates in the measurement model, but such formulation is not considered in this paper.

Logistic models can be used as well to parametrize initial state probabilities and transition probabilities, depending on covariate values. We make use of the following parametrization

$$\log\frac{P(X_0 = s|\boldsymbol{Z}_0)}{P(X_0 = 1|\boldsymbol{Z}_0)} = \beta_{s0} + \boldsymbol{\beta}'_s\boldsymbol{Z}_0, \tag{4}$$

with $1 < s \leq S$, for the initial state probability, and

$$\log\frac{P(X_t = s|X_{t-1} = r, \boldsymbol{Z}_t)}{P(X_t = 1|X_{t-1} = r, \boldsymbol{Z}_t)} = \gamma_{0s} + \gamma_{0rs} + \boldsymbol{\gamma}'_s\boldsymbol{Z}_t, \tag{5}$$

with $1 < s \leq S$, and $1 < r \leq S$ for the transitions probabilities. The first category is taken as reference and the related parameters are set to zero: for the transition model, this means that all parameters related to the elements in the first row and column of the transition matrix are set to zero. Hence, all category-specific effects are to be interpreted in terms of difference from the reference latent state.

The number of free parameters to be estimated is $J \times S$ in the measurement model, $(S-1) \times (K+1)$ in the initial state model, and $(S-1) \times (1 + (S-1) + K)$ in the transition model, where J is the number of response variables, and K the number of covariates. Parameter estimation is normally carried out by Maximum Likelihood.

Let us assume a sample of n independent observations is observed for $T+1$ time occasions, having response vectors $\boldsymbol{y}_{1t}, \dots, \boldsymbol{y}_{nt}$, each being collected along with a vector of covariates, $\boldsymbol{Z}_{1t}, \dots, \boldsymbol{Z}_{nt}$ respectively, with $t = 0, \dots, T$. Under the model specification this paper focuses on, the log-likelihood function is

$$L = \sum_{i=1}^{n} \log P(\boldsymbol{Y}_i = \boldsymbol{y}_i | \boldsymbol{Z}_i). \tag{6}$$

Equation (6) can be maximized by means of the EM algorithm (Dempster et al., 1977). This is an iterative procedure for Maximum Likelihood estimation in the presence of latent variables or missing data. The E-step computes the expected value of the *complete* log-likelihood, which involves obtaining the posterior probabilities for the unknown class memberships, and the M-step maximizes this log-likelihood with respect to the model parameters. However, when using a standard EM algorithm, time and storage required for parameter estimation of LM models increases exponentially with the number of time points (Vermunt et al., 1999). Hence a special version of the EM algorithm is commonly implemented, i.e. the forward–backward algorithm (Baum, Petrie, Soules, & Weiss, 1970; Welch, 2003), in which the size of the problem increases only linearly with the number of time occasions (MacDonald, & Zucchini, 1997). Yet, also estimation with forward–backward algorithm becomes infeasible when the number of time points, the number of indicators and/or the number of covariates one wants to include in the analysis are large. Our three-step methodology proposes an alternative which does not suffer from the issue of dimensionality typical for LM models, yielding estimates that are showed to be correct in finite samples.

Three-step Estimation of LM Models

In this section we describe a three-step estimation approach for LM models and derive the correction to be used in the third step. In the spirit of the work by Asparouhov & Muthén (2014), Bakk et al. (2013), Bolck et al. (2004), and Vermunt (2010) for LC models, the procedure involves 1) estimating a standard LC model without covariates, 2) assigning subjects to latent classes, and 3) analyzing a standard single-indicator latent Markov model with covariates through logistic specifications.

Measurement Model Estimation in Step One

In the first step, the repeated responses from the same individual are treated as being independent, and therefore as corresponding to different sample units. The assumed association structure is depicted in Figure 2. The LC model estimated using the pooled observations is defined as

$$P(\boldsymbol{Y} = \boldsymbol{y}) = \sum_{s=1}^{S} P(X = s) P(\boldsymbol{Y} = \boldsymbol{y} | X = s). \tag{7}$$

Under local independence of responses given class membership, $P(\boldsymbol{Y} = \boldsymbol{y}|X = s)$ can be written as

$$P(\boldsymbol{Y} = \boldsymbol{y}|X = s) = \prod_{j=1}^{J} P(Y_j = y_j | X = s). \tag{8}$$

The model parameters of interest are class proportions $P(X = s)$ and class-specific conditional response probabilities $P(Y_j = y_j|X = s)$. These parameters are estimated maximizing the following log-likelihood function

$$L_{\text{STEP1}} = \sum_{i=1}^{n} \sum_{t=0}^{T} \log P(\boldsymbol{Y}_{it} = \boldsymbol{y}_{it}). \tag{9}$$

Furthermore, as an alternative first step, we also consider parametrizing $P(X = s)$ through a logistic regression, where time variation in the class sizes is taken into account using time dummies[1].

[1]The first authors advocating the usage of time dummies in a panel data logistic regression were Beck, Katz, & Tucker (1998). Carter, & Signorino (2010) point out that using time dummies in a logistic regression causes

Estimation of class membership and classification error in step two

The posterior probability, or posterior class probability (Goodman, 1974a, 1974b, 2007; Hagenaars, 1990; McLachlan & Peel, 2000) $P(X_t = s_t|\boldsymbol{Y}_t = \boldsymbol{y}_t)$, can be expressed, applying Bayes's theorem, as

$$P(X_t = s_t|\boldsymbol{Y}_t = \boldsymbol{y}_t) = \frac{P(X_t = s_t)P(\boldsymbol{Y}_t = \boldsymbol{y}_t|X_t = s_t)}{P(\boldsymbol{Y}_t = \boldsymbol{y}_t)}. \tag{10}$$

Starting from equation (10), different class assignment rules can be applied, the most common of which are modal and proportional assignment. Let us denote the predicted/assigned class by W_t and assume it only depends on $\boldsymbol{Y}_t$, as depicted in Figure 3. Modal assignment estimates W_t allocating a weight $w_{yts} = P(W_t = s_t|\boldsymbol{Y}_t = \boldsymbol{y}_t) = 1$ if $P(X_t = s_t|\boldsymbol{Y}_t = \boldsymbol{y}_t)$ is the largest, and a zero weight otherwise. Whereas modal assignment yields a hard partitioning of the data, proportional assignment treats subjects as belonging to latent class s_t with probability $P(X_t = s_t|\boldsymbol{Y}_t = \boldsymbol{y}_t)$, that is allocating weights $w_{yts} = P(W_t = s_t|\boldsymbol{Y}_t = \boldsymbol{y}_t) = P(X_t = s_t|\boldsymbol{Y}_t = \boldsymbol{y}_t)$. The latter rule yields a soft (or crisp) partitioning of the data. Other assignment rules are available in the literature, such as random assignment (see, for instance, Goodman, 2007). Even if, in principle, both modal and proportional assignment rules can be used, here we focus on modal assignment only. The reason is that proportional assignment may become infeasible with larger number of time points, since each observation has nonzero weights for all possible latent state patterns.

Classification error can be evaluated through the conditional probability of the estimated class value conditional on the true one, that is, $P(W_t = r_t|X_t = s_t)$, where $r_t = 1, \ldots, S$. This probability can be obtained in the following manner

$$\begin{aligned} P(W_t = r_t|X_t = s_t) &= \sum_{\boldsymbol{y}_t} P(W_t = r_t, \boldsymbol{Y}_t = \boldsymbol{y}_t|X_t = s_t) \\ &= \sum_{\boldsymbol{y}_t} P(\boldsymbol{Y}_t = \boldsymbol{y}_t|X_t = s_t)P(W_t = r_t|\boldsymbol{Y}_t = \boldsymbol{y}_t) \\ &= \frac{\sum_{\boldsymbol{y}_t} P(\boldsymbol{Y}_t = \boldsymbol{y}_t)P(X_t = s_t|\boldsymbol{Y}_t = \boldsymbol{y}_t)P(W_t = r_t|\boldsymbol{Y}_t = \boldsymbol{y}_t)}{P(X_t = s_t)}. \end{aligned} \tag{11}$$

issues of quasi-complete separation of the data and inefficiency. As an alternative, they propose to use splines or polynomials of time up to the third order.

As can be seen, the overall classification error probabilities are obtained by averaging over all possible response configurations. However, when the number of response patterns is very large, it is more convenient, as stated in Vermunt (2010), to use the sample patterns. Hence, substituting $P(\boldsymbol{Y}_t = \boldsymbol{y}_t)$ with its empirical distribution yields

$$P(W_t = r_t | X_t = s_t) = \frac{\frac{1}{n(T+1)} \sum_{i=1}^{n} \sum_{t=0}^{T} P(X_t = s_t | \boldsymbol{Y}_{it} = \boldsymbol{y}_{it}) w_{itr}}{P(X_t = s_t)}. \tag{12}$$

The quantity in Equation (12) is strongly related to separation between classes (Vermunt, 2010; Bakk et al., 2013). That is, how well $\boldsymbol{Y}$ predicts the latent classes. Lower separation will cause larger classification error. Measuring class separation - hence classification error - can be done by quantifying how much posterior probabilities deviate from uniform, through the principle of entropy $-\sum_{s_t} P(X_t = s_t | \boldsymbol{Y}_t = \boldsymbol{y}_t) \log P(X_t = s_t | \boldsymbol{Y}_t = \boldsymbol{y}_t)$. An entropy-based (or pseudo) R-squared (Bakk et al, 2013; Magidson, 1981; Vermunt & Magidson, 2005; Vermunt, 2010) is defined as the proportional reduction of entropy when $\boldsymbol{Y}_t$ is available compared to the case in which $\boldsymbol{Y}_t$ is unknown. Such measure yields values ranging from zero (poor classification) to one (perfect classification), and is the reference measure for class separation which will be used in this paper.

A three-step ML method for LM models

In step three we consider the estimated state memberships to be available for each time point. The correction methods proposed for three-step LC analysis (Bolck et al., 2004, and Vermunt, 2010) start from the observation that in an unadjusted three-step estimation procedure, one analyzes the relationship between W_t and $\boldsymbol{Z}_t$, while the relationship of interest is the one between X_t and $\boldsymbol{Z}_t$. This causes underestimation of the covariate effects.

For an adjusted stepwise analysis for LM models, we need to derive the relationship between $P(\boldsymbol{W}|\boldsymbol{Z})$ and $P(\boldsymbol{X}|\boldsymbol{Z})$, were $\boldsymbol{W}$ and $\boldsymbol{X}$ are now vectors. For this derivation, we start from the joint distribution of $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{W}$ given $\boldsymbol{Z}$, which can be defined as follows:

$$P(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{W} | \boldsymbol{Z}) = P(W_0 | \boldsymbol{Y}_0) \ldots P(W_T | \boldsymbol{Y}_T) P(\boldsymbol{Y}_0 | X_0) \ldots P(\boldsymbol{Y}_T | X_T) P(X_0, \ldots, X_T, \boldsymbol{Z}_0, \ldots, \boldsymbol{Z}_T). \tag{13}$$

As can be seen, we use the assumption that $\boldsymbol{Y}$ depends on $\boldsymbol{Z}$ only through $\boldsymbol{X}$, and that classifications obtained in the second step only depend on $\boldsymbol{Y}$. From Equation (13) it is possible to derive the conditional distribution of $\boldsymbol{W}$ given $\boldsymbol{Z}$ by marginalizing over $\boldsymbol{X}$ and $\boldsymbol{Y}$, that is

$$\begin{aligned}
P(\boldsymbol{W}|\boldsymbol{Z}) &= \sum_{X_0,\ldots,X_T}\sum_{\boldsymbol{Y}_0,\ldots,\boldsymbol{Y}_T} P(W_0|\boldsymbol{Y}_0)\ldots P(W_T|\boldsymbol{Y}_T)P(\boldsymbol{Y}_0|X_0)\ldots P(\boldsymbol{Y}_T|X_T)P(X_0|\boldsymbol{Z}_0)P(X_1|X_0,\boldsymbol{Z}_0)\ldots P(X_T|X_{T-1},\boldsymbol{Z}_T)\\
&= \sum_{X_0,\ldots,X_T}\sum_{\boldsymbol{Y}_0,\ldots,\boldsymbol{Y}_T} P(\boldsymbol{Y}_0|X_0)P(W_0|\boldsymbol{Y}_0)\ldots P(\boldsymbol{Y}_T|X_T)P(W_T|\boldsymbol{Y}_T)\underbrace{P(X_0|\boldsymbol{Z}_0)P(X_1|X_0,\boldsymbol{Z}_0)\ldots P(X_T|X_{T-1},\boldsymbol{Z}_T)}_{=\text{MC}}\\
&= \sum_{X_0,\ldots,X_T}\frac{\sum_{\boldsymbol{Y}_0}P(X_0|\boldsymbol{Y}_0)P(\boldsymbol{Y}_0)P(W_0|\boldsymbol{Y}_0)}{P(X_0)}\cdots\frac{\sum_{\boldsymbol{Y}_T}P(X_T|\boldsymbol{Y}_T)P(\boldsymbol{Y}_T)P(W_T|\boldsymbol{Y}_T)}{P(X_T)}\times\text{MC}\\
&= \sum_{X_0,\ldots,X_T}P(W_0|X_0)\ldots P(W_T|X_T)\times\text{MC},
\end{aligned}$$

where the last equality follows from Equation (12), and MC stands for the Markov Chain probabilities of interest. Rearranging terms yields

$$P(\boldsymbol{W}|\boldsymbol{Z}) = \sum_{X_0,\ldots,X_T}P(X_0|\boldsymbol{Z}_0)\prod_{t=1}^{T}P(X_t|X_{t-1},\boldsymbol{Z}_t)\prod_{t=0}^{T}P(W_t|X_t). \tag{14}$$

This shows that the relationship between covariates and the true state memberships X_t can be estimated by defining a latent Markov model with the class assignments W_t as a single indicator, while treating the classification error probabilities $P(W_t|X_t)$ as known. Thus, estimation of the parameters of the structural part of (1) requires maximizing the following log-likelihood function:

$$L_{\text{STEP3}} = \textstyle\sum_{i=1}^{n}\log\sum_{s_0,\ldots,s_T}P(X_0=s_0|\boldsymbol{Z}_{i0})\prod_{t=1}^{T}P(X_t=s_t|X_{t-1}=s_{t-1},\boldsymbol{Z}_{it})\prod_{t=0}^{T}P(W_{it}=r_{it}|X_t=s_t). \tag{15}$$

This is done through a full Baum–Welch algorithm, which overcomes the curse of dimensionality of the standard EM algorithm, while making use of the complete information available from the longitudinal data. At the same time, our procedure eliminates the inconsistency and bias caused by classification errors by means of a ML correction procedure.

The method outlined above can be summarized in the following steps (see also Figure 5).

- **Step one**. A simple LC model is fitted to the pooled data (for each sample unit and time point).

- **Step two**. Posterior membership probabilities are computed, which are used to assign observations to the modal state at each time point. Moreover, the associated classification error probabilities are computed.

- **Step three**. A simple latent Markov model with a single indicator and known error probabilities, which includes the covariates of interest, is estimated. This provides consistent estimates not only of initial and transition probabilities, but also of the regression coefficients in (4) and (5).

Simulation Study

Design

A simulation study was conducted to assess the quality of the proposed three-step methods in latent Markov models with covariates. The three-step method with ML correction was compared with full-information ML (FIML), unadjusted three-step, and three-step with ML correction with the use of time dummies in the first step. The target measures used for the comparisons were bias, standard errors, standard deviations, relative efficiency, and coverage rates. A method performs well when it yields unbiased estimates with small variability, and whenever such variability is correctly retrieved by the SE estimates.

Previous studies on three-step LCA (Bolck et al, 2004; Vermunt, 2010; Bakk et al., 2013) found that class separation - which is strongly related to the amount of classification error as explained earlier - and sample size affect the performance of the step-three correction methods. Class separation can be altered manipulating the class-item associations, the number of items, the number of item categories, and the class sizes. In addition, the presence of covariates will increase class separation, and somewhat more when these covariates are time-varying. In this simulation study, we directly manipulate separation only through the strength of the relationship between items and latent classes. In addition, we consider both time-constant and time-varying covariates.

The data is generated from a three-state latent Markov population model with six dichotomous (low/high) response variables and two numeric covariates, where one is included in both the initial state and the transition model, while the other is only included in the transition model. The first scenario considers time-constant covariates with two and five categories, scored respectively -0.5, 0.5 and -2, -1, 0, 1, 2. The second scenario deals with time-varying covariates, generated from two independent AR(1) processes, both with standard normal white noise. Notice that in both scenarios, covariates have means of zero (Table 1).

The classes' profiles are set such that Class 1 is likely to give a high response on items 4 and 6, Class 2 on the first three items, and Class 3 on the first two and the last two items (see Table 2). The response probabilities for the most likely responses were set to 0.8 and 0.9, corresponding to moderate and high class separation in the first step model. These settings correspond to entropy-based R^2 of, respectively, 0.64 (moderate separation) and 0.9 (high separation) with time-constant covariates, and of respectively 0.6 (moderate separation) and 0.85 (high separation) with time-varying covariates. The sample sizes used were 100, 500 and 1,000. The number of time points was fixed to five.

Covariates Z_1 and Z_2 enter the structural models through a multinomial logit parametrization as shown in equations (4) and (5), with the first class as reference category. The intercepts are such that stability is larger than change, as is typically the case in latent Markov models. We set the regression coefficients such that a higher value of Z_1 corresponds with lower initial state probabilities for and less transitions towards latent states 2 and 3. In contrast, higher values of Z_2 correspond with a tendency to move out of class 1. More specifically, the intercepts of the initial state model where both set to 0, while $\beta_2 = \beta_3 = -0.5$. The transition model intercept terms are set to $\gamma_{02} = \gamma_{03} = -2$, $\gamma_{022} = \gamma_{033} = 4$ and $\gamma_{023} = \gamma_{032} = 2$. Covariate parameters in the logit model for the transition probabilities were set, respectively, both to -1 and to 0.25.

Furthermore, we consider cases where state proportions are unequal at the initial time point. We model this through intercept values ranging from -0.1 (slightly unequal) to -2.5 (strongly unequal). Given a symmetric transition matrix, this implies a larger time variation in the state distribution, as more persons will tend to move from state one to other states (for a sufficiently large time span, this would cause states to become eventually equal in size).

For all 96 simulation conditions, obtained by combining the specified class separation levels, sample sizes, and covariate types, we used 500 replications and estimated the model using the 4 methods being compared. Data generation and parameter estimation were carried out using Latent GOLD version 5.0 (Vermunt and Magidson, 2013).

Simulation results

Table 3 presents bias, average standard errors (SEs), standard deviations (SDs), and coverage of 95% CI's (CVG) for all four methods for the condition with time-constant covariates, averaged over sample size and separation conditions. As can be seen, standard 3–step has the largest overall bias, where the encountered underestimation of the covariate effects is stronger in the model for the transitions than in the model for the initial states. However, as will be shown below, bias varies strongly across separation levels and sample sizes. For both corrected 3-step methods (with and without time dummies in the first step) we found similar (small) bias, which moreover is close to the bias levels encountered using the one-step FIML approach.

Estimated SDs across all conditions show that both correction methods are similar to FIML in terms of overall efficiency (above 0.9). Comparing SEs to SDs shows that the average ratios are slightly below 1, indicating that overall SEs tend to be slightly underestimated. As shown below, underestimation vanishes as class separation and sample size become larger.

An approximate 95% coverage is achieved in the two corrected stepwise routines, which indicates that confidence intervals have coverage equal to nominal coverage levels. The values are close to those for the one-step FIML method and indicate that SEs correctly display the actual variability of the estimates - together with unbiasedness of the estimators.

For the sake of brevity, we omit results for time-varying covariates, as we see the same patterns as with time-constant covariates. The only difference is that the variability of the estimates (SDs) is about 50% lower, which is probably caused by the fact that an $AR(1)$ covariate process provides additional information on the covariate parameter values. Hence, below we will focus on results on time-contant covariates only, as they are informative also for the time-varying covariate case.[2]

In Figures 6 and 7 we show boxplots of the estimated bias for γ_{12} and γ_{23} for all six conditions with time-constant covariates and for all four estimation methods, across simulated

[2]Tables with results for all parameters under all 96 conditions can be found in the online appendix.

samples. As already noted in the overall bias values presented above, retrieving covariate effects in the transitions model was more challenging. In the standard three-step, class separation has a stronger effect than sample size in improving the accuracy of the estimates. Even with a small sample size (100), increasing class separation shifts the estimated bias towards zero.

The corrected three-step methods' estimates are in line with FIML, except for the combination of small sample size and moderate class separation. This result agrees with what is reported in Bakk et al. (2013) and Vermunt (2010) for three-step LC analysis. In effect, in the small sample size and moderate class separation situation, the differences between classes are somewhat overestimated in the first step, yielding estimated classification errors, and thus corrections, which are too optimistic. We notice also that adding time dummies in the first step increases the variability of the estimates (see conditions three to six) compared to the standard three-step correction. While the idea behind the use of time-dummies is to make the first-step estimation more robust to time variation, as class separation increases, the standard ML-corrected three-step estimator becomes more accurate (hence making time-dummies estimator relatively inefficient).

Empirical Example: Household Acquisition of Financial Products

In this section we illustrate the use of the proposed three-step method using an empirical example using a data set that was analyzed earlier by Paas, Bijmolt, & Vermunt (2007), and Pass, Vermunt, & Bijmolt (2007). The data are collected by the Dutch division of the international market research company Growth from Knowledge through face-to-face interviews. Respondents were asked about their ownership of 12 financial products in 1996, 1998, 2000 and 2002. Also information on net household income, age of the head of the household, and size of the household is available. The study includes 7676 households. Due to either attrition or late sign-up, some of them did not participate in all waves of the panel. Replacement of households dropping out is conducted in such a way that the sample remains representative for the population with respect to several important demographic variables, such as age, income, and marital status. This is why the missingness at random assumption seems to be appropriate for the present analysis.

Pass, Vermunt, & Bijmolt (2007) used a model specification with a time-constant measurement model and time homogeneity in the transitions, which is the same working assumption we use here.[3] As they do not include time as auxiliary variable, we used the three-step specification without time dummies in the first step. The number of classes, or segments, was selected using BIC computed using the model likelihood from the first step. BIC selected 9 latent classes, which can be interpreted as prototypical product portfolios. This is consistent with the minimum-BIC class number in Paas et al. (2007).

Table 4 shows the results of the first-step model (measurement part). The results suggest an order of acquisition of financial products consistent with Paas et al. (2007) and references therein. Basic financial products - e.g. savings account and car insurance - have the highest penetration. Latent classes showing a high probability of owning the more risky products - like bonds and shares - tend to own the more basic financial products. This general pattern suggests a common order of acquisition. On the other hand, the model shows some deviations from the common order. Being able to elicit a common order of acquisitions, together with possible deviations, is a strength of this kind of analysis. As an example, consider the penetration rate of credit card in class 6. This is higher than mortgage or house insurance penetration rates, whereas these two products have a higher overall penetration.

The estimated transition probabilities from the third-step model of remaining in the same segment are close to 1, ranging from 0.87 up to 0.99.[4] It is very likely for a household to stay either in one of the first two classes - the ones with high penetration products - or in one of the last three - where portfolios include also products with lower penetration. However, we found that 27 % of the 7676 households switched segment at least once during the time span 1996-2002. This percentage is consistent with the literature on household finance and expected gradual development of a household's financial product portfolios (Guiso, Haliossos, & Jappelli, 2002). Most of changes concern changes towards segments with a more diversified portfolio. Segment 3 registers a 6 % probability to switch to segment 6 and segment 8.

[3]Because people change their portfolio composition very slowly, we smooth the estimated probabilities slightly away from zero by imposing priors on the parameters of the logistic models for the transition probabilities (see Galindo-Garre, Vermunt, & Bergsma, 2004).

[4]The full table of estimated transition probabilities can be found in the online appendix.

Segments 5 and 6 have nonzero probability of switching to higher segments, and segment 8 shows 6 % probability to switch to segment 9 in the next period. This points out a gradual development of households portfolios. A household having, say, a segment one portfolio prototype, will step through intermediate states before getting to higher segments, like 8 or 9.

The step-three model also contained covariates, whose effects are reported in Table 5. Younger heads of household are more likely to start in segment 1, 3, or 4. Higher income and older age of the head of the household increase the likelihood to switch from lower segments to intermediate segments like 5 and 6. In contrast, household size tends to negatively correlate with the probability of being in intermediate segments, with exception of segment 6. Here the higher incomes seem to compensate for higher household expenses due to household size. Households with relatively older heads of household and many members are likely to expand their portfolios from intermediate segments to segments with relatively high ownership probabilities for most products.

The complexity of one-step analysis lies on the fact that one has to select the model in terms of number of latent classes, whereas the model needs to be re-estimated until the one minimizing BIC is found (like in this example). In addition, model selection includes also the decision about which covariate to include in the analysis. Letting an additional covariate in (or leaving it out) requires as well re-estimation of the full model. Although this is possible, it is very cumbersome. While the results we obtained are similar to those reported by Paas et al. (2007), our stepwise estimation approach turned out to be much more practical. Note the simultaneous model by Paas et al. (2007) was huge given the 7676 observations, 4 time points, 12 indicators, and several covariates, especially given the large number of latent states needed.

In a three-step approach, selection of the number of latent classes is very easy (and quick), as it is done using a simple latent class model on the pooled observations in step one. Once the latent class structure is derived (step one), selecting the covariates to be included in the analysis and studying their effect on class memberships is done with the step-three model. This allowed us to easily deal with the complexities (including parameter smoothing) of a big data set like the one used in Paas et al. (2007).

Discussion

The present paper showed how to make LM modeling with potentially many responses and covariates per time point feasible, using a bias-adjusted three-step procedure. We have worked out the theoretical foundation of Vermunt (2010)'s ML correction for latent Markov models. In addition, we have provided a time-robust alternative first-step model. Our extended simulation studies showed the conditions under which the procedure is trustworthy, complementing the work by Asparouhov and Muthén (2014).

The results of the simulations indicate, as expected, that a standard three-step leads to downward bias in the estimation of covariate effects. The correction methods perform well in terms of parameter estimates and related SEs, except for cases with small sample size ($n < 500$) and bad class separation ($R^2_{\text{ENTR}} < 0.6$). We found a slight loss in efficiency of the first-step robust three-step ML method relative to normal three-step ML, as sample size becomes larger and classes well separated. With slightly lower class separation, we recommend the use of a time-robust first step. In all other cases, the more efficient three-step ML should be used.

In the empirical example we show that our method is able to easily deal with 30704 (7676×4) individual records on 12 items, plus 3 individual and time-specific covariates, in contrast with the standard method. This is not surprising, as the step-three model is a single-indicator latent Markov model, for which computational complexity increases only linearly with the number of time points, and with the number of covariates per time point. On the other hand, computational complexity is left unaffected by an increasing number of response variables. This makes the three-step method suitable especially for larger-scale problems.

Asparouhov and Muthén (2014) also considered the situations with different measurement models per time point. Our three-step method can easily be adapted to allow for different measurement models per time point as follows. Step one and two of the analysis should then be carried out separately for each of the time points. In the third step, one should account for the fact that the classification error probabilities are time-specific. For an example, see Nylund-Gibson et al. (2014) who even combined two different classification models: a LC measurement model and a mixture growth measurement model to assess how kindergarten readiness profiles link to elementary students' trajectories.

The current study has various limitations. First, we studied the behavior of the bias-adjusted three-step approach only for the situation in which model assumptions hold - for example, that covariates do not affect responses directly. Future research could focus on the behavior of and possible adaptation of the three-step method to situations in which assumptions are violated. Second, as indicated by Bakk et al. (2013), Bakk, Oberski, & Vermunt (2014), and Vermunt (2010), in the step-three adjustments the classification error probabilities are treated as known, whereas in fact these are computed using the parameters estimated in the step-one model. This treatment of estimates as fixed values may cause underestimation of standard errors in the third step (Bakk et al., 2014). If uncertainty about fixed parameters is higher (low separation between classes and small sample size), inference with uncorrected SEs may no longer be valid. Bakk et al. (2014) showed how to obtain asymptotic standard errors of the step-three parameters, taking into account that the step-three model uses quantities based on the step-one model estimates (rather than on their population values). While the same approach can be also used in three-step LM modeling, the step-one SEs would have also to take into account that observations are not independent. Third, it is not fully clear how proper step-one model selection has to be carried out taking into account the dependence between the observations. Finding a suitable information criterion, with a penalty term that takes the time structure into account, and comparing it with standard measures is an interesting topic for future research. Fourth, in the second step we opted for the more convenient modal assignment, whereas proportional assignment may increase the accuracy of the method. However, with proportional assignment, implementation of a full Baum-Welch algorithm for the third-step model seemed to be infeasible, as all possible state assignment patterns would have to be processed for each observation. An efficient algorithm allowing for proportional assignment hence requires further research.

References

Asparouhov, T., & Muthén, B. (2014). Auxiliary variables in mixture modeling: Three-step approaches using Mplus. *Structural Equation Modeling: A Multidisciplinary Journal, 21*(3), 329–341.

Bakk, Z., Oberski, D.L. & Vermunt, J.K. (2014). Relating latent class assignments to external variables: standard errors for corrected inference. *Political Analysis*, 22, 520–540.

Bakk, Z., Tekle, F.B. & Vermunt, J.K. (2013). Estimating the association between latent class membership and external variables using bias-adjusted three-step approaches. *Sociological Methodology, 43*(1), 272–311.

Bartolucci, F., Montanari, G. & Pandolfi, S. (2015). Three-step estimation of latent Markov models with covariates. *Computational Statistics and Data Analysis, 83*, 287–301.

Baum, L.E., Petrie, T., Soules, G. & Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*, 164–171.

Beck, N., Katz, J. N. & Tucker, R. (1998). Taking time seriously: Time-series-cross-section analysis with a binary dependent variable. *American Journal of Political Science*, 1260–1288.

Bolck, A., Croon, M.A. & Hagenaars, J.A. (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political Analysis, 12*(1), 3–27.

Carter, D. B. & Signorino, C. S. (2010). Back to the future: Modeling time dependence in binary data. *Political Analysis,* mpq013.

Collins, L.M., & Lanza, S. T. (2010). *Latent class and latent transition analysis*. Hoboken, NJ: Wiley.

Collins, L. M., & Wugalter, S. E. (1992). Latent class models for stagesequential dynamic latent variables. *Multivariate Behavioral Research*, 27, 131–157.

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

Galindo–Garre, F., Vermunt, J.K., & W. Bergsma (2004). Bayesian posterior estimation of logit parameters with small samples. *Sociological Methods and Research, 33*(1), 88–117.

Goodman, L. A. (2007). On the assignment of individuals to latent classes. *Sociological Methodology, 37*(1), 1–22.

Goodman, L.A. (1974a). The analysis of systems of qualitative variables when some of the variables are unobservable. Part IA modified latent structure approach. *American Journal of Sociology*, 1179-1259.

Goodman, L.A. (1974b). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika, 61*(2), 215–231.

Guiso, L., Haliossos, M., & Jappelli, T. (2002). *Household Portfolios*. MIT Press.

Hagenaars, J. A. (1992). Exemplifying longitudinal loglinear analysis with latent variables. In P G. M. Van der Heijden, W. Jansen, B. Francis, and G. U. H. Seeber (Eds.), *Statistical modelling,* 105–120. Amsterdam: Elsevier Science Publishers B.V.

Lazarsfeld, P.F. (1950). The logical and mathematical foundation of latent structure analysis. *Measurement and prediction, 4.*

Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Boston: Houghton Mill.

MacDonald, I.L., & Zucchini, W. (1997). *Hidden Markov and other models for discrete-valued time series*. Chapman & Hall / CRC Press.

Magidson, J. (1981). Qualitative variance, entropy, and correlation ratios for nominal dependent variables. *Social Science Research, 10*(2), 177-194.

Nylund-Gibson, K., Grimm, R., Quirk, M., & Furlong, M. (2014). A Latent Transition Mixture Model Using the Three-Step Specification. *Structural Equation Modeling: A Multidisciplinary Journal, 21*(3), 439-454.

Paas, L. J., Bijmolt, T. H., & Vermunt, J. K. (2007). Acquisition patterns of financial products: A longitudinal investigation. *Journal of Economic Psychology, 28*(2), 229–241.

Pass, L. J., Vermunt, J. K., & Bijmolt, A. (2007). Discrete time, discrete state latent Markov modeling for assessing and predicting household acquisitions of financial products. *Journal of the Royal Statistical Society: Series A, 170*(4), 955–974.

Van de Pol, F., & De Leeuw, J. A. N. (1986). A latent Markov model to correct for measurement error. *Sociological Methods and Research, 15*(1-2), 118–141.

Van de Pol, F., & Langeheine, R. (1990). Mixed Markov latent class models. *Sociological methodology, 20*, 213–247.

Vermunt, J. K., & Magidson, J. (2013). *Technical Guide for Latent GOLD 5.0: Basic, Advanced, and Syntax.* Statistical Innovations Inc., Belmont.

Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political analysis, 18*(4), 450–469.

Vermunt, J.K., & Magidson, J. (2008). *LG-Syntax User's guide: Manual for Latent GOLD 4.5 Syntax Module.* Statistical Innovations Inc., Belmont.

Vermunt, J.K., & Magidson, J. (2005). *Latent GOLD 4.0 User's guide.* Statistical Innovations Inc., Belmont.

Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. *Applied latent class analysis, 11*, 89–106.

Vermunt, J. K., Langeheine, R., & Bockenholt, U. (1999). Discrete-time discrete-state latent Markov models with time-constant and time-varying covariates. *Journal of Educational and Behavioral Statistics, 24*(2), 179–207.

Vermunt, J. K. (1997). *Log-linear models for event histories.* Advanced Quantitative Techniques in the Social Sciences Series. Thousand Oaks: Sage Publications.

Welch, L. R. (2003). Hidden Markov models and the Baum-Welch algorithm. *IEEE Information Theory Society Newsletter, 53*(4), 10-13.

Wiggings, L.M. (1973). *Panel analysis: Latent probability models for attitude and behavior processes.* Elsevier Scientific, San Francisco.

Table 1
Types of covariates used in the simulation.

	Time-constant	Time-varying	Initial pr.	Transitions
Z_1	(-0.5,0.5)	$AR(1)$	✓	✓
Z_2	(-2, -1, 0, 1, 2)	$AR(1)$	X	✓

Table 2
Item-class probabilities in the simulation.

Item	Class 1	Class 2	Class 3
1	low	high	high
2	low	high	high
3	low	high	low
4	high	low	low
5	low	low	high
6	high	low	high

Table 3

Simulation Results - Parameters bias, SE, SD, SE/SD and coverage rates (CVG) averaged over all simulation conditions. Time-constant covariates.

	$\boldsymbol{\beta} = (-0.5, -0.5)'$									
	Bias		SE		SD		SE/SD		CVG	
FIML	-0.005	-0.011	0.319	0.325	0.328	0.339	0.982	0.963	0.945	0.941
3S naive	0.047	0.057	0.294	0.296	0.297	0.308	0.997	0.967	0.939	0.927
3S ML	0.003	-0.005	0.331	0.345	0.343	0.359	0.977	0.967	0.947	0.941
3Sdum	-0.002	-0.008	0.337	0.348	0.352	0.360	0.971	0.968	0.946	0.946
	$\boldsymbol{\gamma}_{Z1} = (-1, -1)'$									
	Bias		SE		SD		SE/SD		CVG	
FIML	-0.021	-0.023	0.225	0.229	0.230	0.234	0.990	0.992	0.950	0.947
3S naive	0.140	0.161	0.181	0.176	0.187	0.173	0.965	1.009	0.735	0.704
3S ML	-0.015	-0.007	0.233	0.239	0.238	0.242	0.992	0.996	0.951	0.950
3Sdum	-0.011	-0.004	0.232	0.236	0.238	0.238	0.989	1.001	0.949	0.950
	$\boldsymbol{\gamma}_{Z2} = (0.25, 0.25)'$									
	Bias		SE		SD		SE/SD		CVG	
FIML	0.003	0.003	0.079	0.080	0.080	0.084	1.001	0.965	0.952	0.947
3S naive	-0.050	-0.058	0.064	0.062	0.064	0.063	0.986	0.977	0.754	0.717
3S ML	-0.001	-0.001	0.081	0.083	0.081	0.087	1.012	0.961	0.952	0.942
3Sdum	-0.002	-0.002	0.081	0.082	0.081	0.086	1.011	0.961	0.954	0.943

Table 4
Measurement model estimates (conditional probabilities) in the example analysis.

Product	Latent class								
	1	2	3	4	5	6	7	8	9
Bonds	0.00	0.03	0.00	0.01	0.34	0.03	0.07	0.00	0.11
Shares	0.00	0.04	0.00	0.04	0.70	0.15	0.15	0.01	0.38
Investment trust	0.01	0.16	0.03	0.06	0.57	0.39	0.31	0.10	0.60
Unemployment insurance	0.00	0.01	0.15	0.04	0.02	0.36	0.00	0.34	0.46
Loan	0.12	0.03	0.45	0.18	0.08	0.35	0.00	0.31	0.25
Credit card	0.05	0.13	0.29	0.31	0.45	0.56	0.41	0.43	0.75
Mortgage	0.01	0.00	0.00	0.92	0.14	0.02	0.73	0.99	0.93
House insurance	0.03	0.34	0.02	0.93	0.61	0.06	0.99	0.99	0.98
Life insurance	0.18	0.14	0.68	0.64	0.35	0.65	0.41	0.96	0.95
Pension fund	0.25	0.28	0.85	0.38	0.24	0.93	0.49	0.95	0.94
Car insurance	0.29	0.70	0.82	0.65	0.76	0.88	1.00	0.93	0.93
Savings account	0.77	0.97	0.92	0.88	0.89	1.00	0.98	0.99	0.99
Average number of products	1.71	2.83	4.22	5.03	5.14	5.37	5.55	7.01	8.26

Table 5

*Estimates of regression parameters in transition probability model. Covariates are age (*age*), household size (*hsize*), and income (*income*). Standard errors in parenthesis.*

Latent Class	γ_{age}	γ_{hsize}	γ_{income}
1	-0.52	-1.10	0.71
	(0.16)	(0.29)	(0.22)
2	1.02	0.90	-0.40
	(0.18)	(0.18)	(0.21)
3	-0.74	0.75	-3.28
	(0.19)	(0.36)	(0.71)
4	-0.57	-0.87	0.30
	(0.27)	(0.41)	(0.50)
5	0.14	-0.07	0.12
	(0.14)	(0.16)	(0.17)
6	0.65	0.39	2.82
	(0.29)	(0.22)	(0.45)
7	-0.91	-0.35	0.36
	(0.38)	(0.40)	(0.41)
8	0.58	0.27	-0.53
	(0.12)	(0.15)	(0.18)
9	0.35	0.09	-0.1
	(0.16)	(0.21)	(0.23)

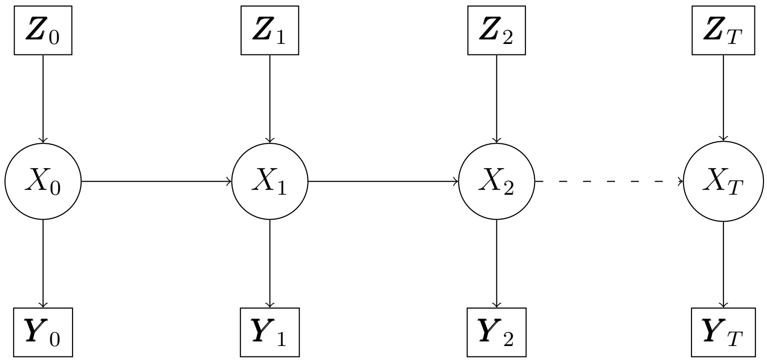

Figure 1. Latent Markov (LM) model graph

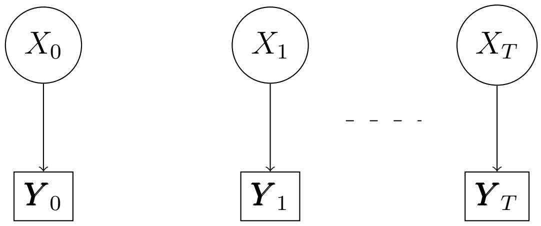

Figure 2. Step 1: Latent Class model.

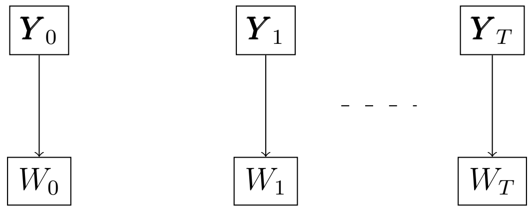

Figure 3. Step two: Assignments and classification error

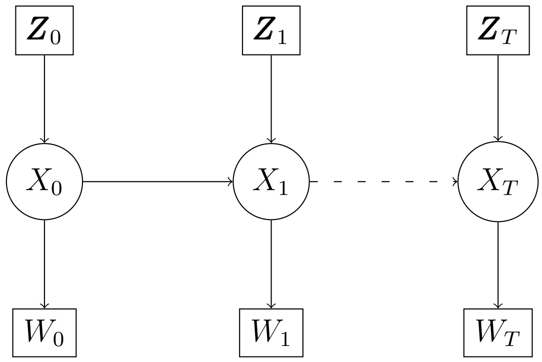

Figure 4. Step three: latent Markov model with single indicators.

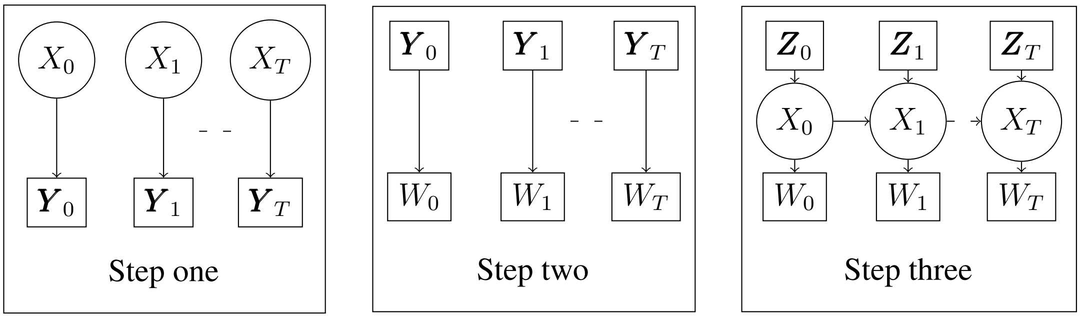

Figure 5. Summary of the three steps

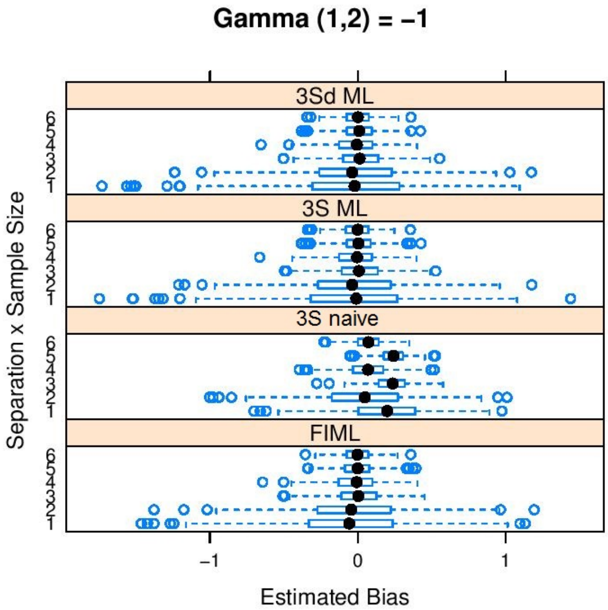

Figure 6. Boxplot of estimated bias across simulated samples, conditional to separation–sample size combination (from 1=moderate separation and small sample size, to 6=high separation and large sample size) and method (3Sd ML = 3-step with ML correction and time dummies in the first step; 3S ML = 3-step with ML correction; 3S naive = 3-step without correction; FIML = Full Information Maximum Likelihood).

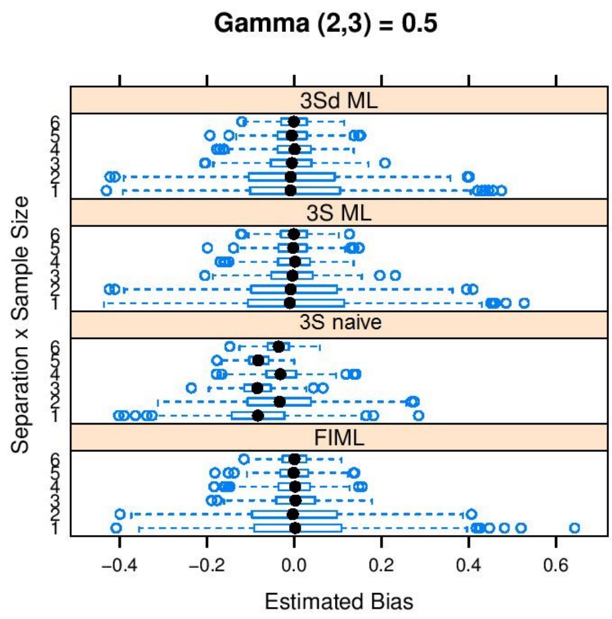

Figure 7. Boxplot of estimated bias across simulated samples, conditional to separation–sample size combination (from 1=moderate separation and small sample size, to 6=high separation and large sample size) and method (3Sd ML = 3-step with ML correction and time dummies in the first step; 3S ML = 3-step with ML correction; 3S = 3-step without correction; FIML = Full Information Maximum Likelihood).

Appendix

Latent GOLD syntax for the empirical example

Latent GOLD syntax for step one and two

```
options
    output
    parameters=effect betaopts=wl standarderrors profile
    probmeans=posterior bivariateresiduals
    estimatedvalues=model;
    outfile 'C:\Users\classification.sav'
    classification=model keep K, L, M, time, id;
variables
    dependent Y1, Y2, Y3, Y4, Y5, Y6, Y7, Y8, Y9, Y10, Y11,
    Y12;
    latent
    Cluster nominal 9;
equations
    Cluster <- 1;
    Y1-Y12 <- 1 + Cluster;
```

Latent GOLD syntax for step three

This is Latent GOLD syntax for step three with covariates in the initial probability and transition models.

```
options
    step3 modal ml;
    output
    iterationdetails parameters=effect betaopts=wl
    standarderrors profile probmeans=model
    estimatedvalues=model;
variables
```

```
    caseid id;
    independent K, L, M, time nominal;
    latent Cluster nominal dynamic posterior = (Cluster#1
    Cluster#2 Cluster#3 Cluster#4 Cluster#5 Cluster#6
    Cluster#7 Cluster#8 Cluster#9);
equations
    Cluster[=0] <- 1 + K + L + M;
    Cluster <- 1 + Cluster[-1] + K + L + M;
```

In order not to include covariates in the model, it suffices to remove `K`, `L`, and `M` from the equations for the initial probability and transition models. Equations now read

```
    Cluster[=0] <- 1;
    Cluster <- 1 + Cluster[-1];
```