Mostly harmless direct effects: a comparison of different latent Markov modeling approaches

Roberto Di Mari
Department of Economics and Business, University of Catania, Italy

Zsuzsa Bakk
Statistics Unit, Methodology, Institute of Psychology, Leiden University, The Netherlands

Abstract

We evaluate the performance of the most common estimators of Latent Markov (LM) models with covariates in the presence of direct effects of the covariates on the indicators of the LM model. In LM modeling it is common practice not to model such direct effects ignoring the consequences that might have on the overall model fit and the parameters of interest. However, in general literature about latent variable modeling it is well known that unmodeled direct effects can severely bias the parameter estimates of the model at hand. We evaluate how the presence of direct effects influences the bias and efficiency of the three most common estimators of LM models, the one-step, two-step and three-step approaches. Furthermore we propose amendments (that were thus far not used in the context of LM modeling) to the two and three-step approaches that make it possible to account for direct effects and eliminate bias as a consequence. This is done by modeling the (possible) direct effects in the first step of the stepwise estimation procedures. We evaluate the proposed estimators through an extensive simulation study, and illustrate them via a real data application. Our results show, first, that the augmented two-step and three-step approaches are unbiased and efficient estimators of LM models with direct effects. Second, ignoring the direct effects leads to biased estimates with all existing estimators, the one-step approach being the most sensitive.

Keywords: latent Markov model, direct effects, model comparison, stepwise modeling

Address:
Room 20,
Corso Italia, 55, 95129 Catania
Phone: +390957537646
Email: roberto.dimari@unict.it

Introduction

The current paper proposes to give an overview of the most used estimators of Latent Markov (LM) models (one-step, two-step and three-step estimators) in the presence of direct effects (DE) of covariates on the indicators of the LM model. The goal of the paper is mostly educational: to introduce the most popular estimators, describe their performance, and give practical advice on which estimators to use in the presence of DE's. Furthermore, based on our real data example, we provide also a detailed Latent GOLD syntax that serves as a step-by-step guide for applied researchers interested in applying these approaches. Next to showing how the most used estimators perform in the presence of DE, we also propose alternative implementations of the stepwise estimators that can model DE, and that were so far not proposed in literature.

Latent Markov models are widely used in the social and behavioral sciences for modeling change over time. In its basic form, the model includes the repeated measurement over time of one or more variables of interest. Examples include estimating the change in brand choice behavior over time (Poulsen, 1990) or modeling disease progression over time (Jackson et al., 2003).

These models assume the existence of a latent process, typically modeled as a first-order Markov chain with finite state space, which affects the distribution of the response variables. A central assumption behind this approach is that the response variables are conditionally independent given this latent process. This is known as local independence, meaning that the latent process fully explains the observable behavior of a subject.

An extension of the model includes (time constant and time varying) covariates that predict the probability of belonging to certain latent states, or the transition probability between states over time. There are three main approaches to estimate LM models with covariates: one-step, two-step and three-step approaches.

Using the one-step approach (also known as full information maximum likelihood, FIML), the complete model of interest is estimated simultaneously. Although algorithms for simultaneous ML estimation exist (like the Baum-Welch type of EM algorithm), this approach is infeasible, hence hardly ever used in practice, when the number of time points, indicators and/or covariates is large. Furthermore, every time a small change is made to the model (for example adding a direct effect), the full model of interest has to be re-estimated. Because of these difficulties, stepwise estimation procedures, that break down the problem into smaller steps, have been suggested in literature.

One well known alternative is the three-step approach, whereby in step one the measurement model is estimated using a latent class (LC) model on the pooled data, then respondents are assigned to latent states (step 2), and in step three the assigned scores are related to covariates. A problem with this approach is that the third step estimates are biased, due to ignoring the classification error introduced in step two. A bias-adjustment method was developed by J. K. Vermunt (2010) for LC models (ML adjustment method), extended to LC models with distal outcomes by Bakk et al. (2013) (see also Zhu et al., 2017, for an extension to the LC model with distal outcome and multiple latent variables). Later on, Di Mari, Oberski, & Vermunt (2016) implemented the ML adjustment method to LM models, which obtains unbiased estimates as long as the main assumptions are met.

Alternatively a two-step estimator proposed by Bartolucci, Montanari, & Pandolfi (2014) can be applied. Similarly to the three-step approach, in step one a LC model is estimated on the pooled data. In step two a LM model with covariates is estimated, while keeping the measurement model parameters fixed at the estimates obtained in step one. This approach is similar to the well known two-stage least square estimator from the econometric literature (Murphy & Topel, 1985; Hardin, 2002). Using this approach, the complex estimation specific to FIML is broken down to manageable sized problems.

In the current paper we evaluate the bias and efficiency of the one-step, two-step and three-step approaches in the presence of direct effects. While using all of these approaches it is a common assumption that the covariates do not influence directly the indicators, the violation of this assumption was not evaluated thus far in the context of LM modeling. It is well known that both the one-step and three-step approaches are subject to severe bias when direct effects are present in latent class (LC), latent trait (Asparouhov & Muthén, 2014), and regression mixture models (Kim et al., 2016; Nylund-Gibson & Masyn, 2016). Notably, the amount of bias is increasing in the number of direct effects. As such it is of utmost importance to understand the consequences of the presence of unmodeled direct effects for LM models as well. A work which explores the available estimators for LM modeling in the presence of direct effects is lacking, and the present paper fills this gap. In addition, although it is not new in LM modeling the use of the two-step estimator (Bartolucci et al., 2014; Montanari & Pandolfi, 2017), this needs deeper investigation under a broad set of possible conditions, including the presence of unmodeled direct effects.

Furthermore, we propose alternative implementations of the stepwise approaches that model the direct effects in step one. Our proposal is motivated by the findings in Bennink et al. (2015) for multilevel LC models, and Asparouhov & Muthén (2014) for LC models. Both papers show that including the direct effects in the first step leads to unbiased estimates of the parameters of interest in the last step model of LC models. Using the two-step approach, in step one the direct effects can be modeled, and the step two estimates conditioned on this adjusted step one model. Using this implementation, the final model contains the estimated effect of the covariates on the latent states and the transition probabilities, but also the direct effects, making it possible to interpret the direct effects, which is of interest in certain implementations.

Using the three-step approach, we propose an implementation that accounts for possible direct effects by modeling them as noise in an overparametrized step one model. Such an alternative makes it unnecessary to model the correct direct effects exclusively, rather using an augmented model that accounts for multiple possible direct effects. Then, the noise created by misspecifying the direct effects in the first step is dealt with by adding an absorbing second latent variable.

This alternative three-step implementation is needed, because in the three-step approach, contrary to the one-step and the two-step approaches, using residual statistics to identify the correct direct effects cannot be done. In the augmented three-step approach, although it is not possible to interpret the direct effects in step three, it is possible to model them in a way that their presence does not affect the estimates of the parameters of interest.

The paper proceeds as follows: first we introduce the LM model in its most general form, discussing one-step, two-step and three-step ML estimation, including the proposed modifications to the stepwise approaches. Next we evaluate, using a simulation study, the performance of the various approaches and proceed to a real data example. In our conclusion we give a practical guideline to applied researchers about which approach to use under different circumstances.

The latent Markov model with covariates

The LM model, in its baseline formulation, was first introduced by Wiggins (1973). Known also under the name of hidden Markov model (MacDonald & Zucchini, 1997), or latent transition model, was later on extended to include covariates in the initial state and transition probabilities by J. K. Vermunt et al. (1999), and in the measurement model by Bartolucci & Farcomeni (2009).

Extensive reviews on LM can be found in Bartolucci et al. (2012), and Zucchini et al. (2016). We first present the LM model with covariates in the initial and transition probabilities, then consider as well the case where covariates have also a direct effect on the indicators - therefore enter the measurement model - and sketch maximum likelihood one-step estimation.

Let $Y_{1t}, \dots, Y_{Jt}$ be J indicator variables measured for $T+1$ time occasions, with $t = 0, \dots, T$. Let us indicate as $\mathbf{Y}_t$ the full response pattern at time t, with realizations $\mathbf{Y}_t$, observed alongside with K covariates $\mathbf{Z}_t$. In addition, we denote the full set of response configurations and available covariates at the $T+1$ measurement occasions respectively as $\mathbf{Y}$ and $\mathbf{Z}$. The latent variable X_t has a finite state space of size S and we let s_t be one of the S possible latent states for time t. A latent Markov model formulates the probability of having a particular sequence of response configurations at the $T+1$ time occasions, given the time-specific covariates, as follows

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

The model defined in Equation 1 is depicted in Figure 1.

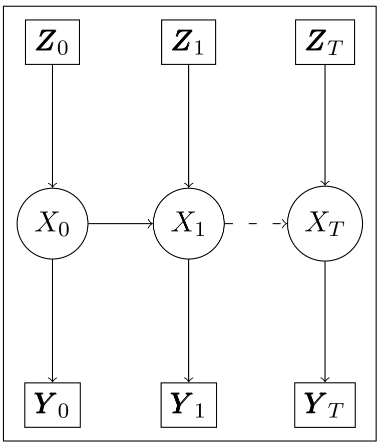

Figure 1. Latent Markov model with covariates: one-step model

For the structural component we assume a first-order time-homogeneous Markov chain. Covariates enter both in the initial state probability $P(\mathbf{X}_0 = s_0|\,\mathbf{Z}_0)$, and the latent transition probability $P(\mathbf{X}_t = s_t|\mathbf{X}_{t-1} = s_{t-1}, \mathbf{Z}_t)$.

The measurement component, as for standard latent class models for dichotomous responses, can be written as

$$P(\mathbf{Y}_t = \mathbf{y}_t|\mathrm{X}_t = s_t) = \prod_{j=1}^{J} P(\mathrm{Y}_{jt} = \mathrm{y}_{jt}|\mathrm{X}_t = s_t) \tag{2}$$

due to the assumption of mutual independence of the response variables given class membership, i.e. local independence. Although we assume time homogeneity, the measurement can be allowed to differ across time, and/or it can include direct effects of the covariates on one or more items. In our case it is natural to assume that, given the latent variable, Y_{jt} has a Bernoulli distribution with success probability π_{sj}.

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 as

$$\log\left(\frac{\pi_{sj}}{1-\pi_{sj}}\right) = \theta_{js}, \quad \text{for } 1 \leq s \leq S, \tag{3}$$

or

$$\log\left(\frac{\pi_{sj}}{1-\pi_{sj}}\right) = \theta_{js} + \boldsymbol{\alpha}_j'\,\mathbf{Z}, \quad \text{for } 1 \leq s \leq S, \tag{4}$$

if direct effects of the covariantes are also included.

We estimate intitial state probabilities and transition probabilities using logistic models with the following parametrization

$$\log\frac{P(X_0 = s|\,\mathbf{Z}_0)}{P(X_0 = 1|\,\mathbf{Z}_0)} = \beta_{s0} + \boldsymbol{\beta}_s'\,\mathbf{Z}_0, \tag{5}$$

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

$$\log\frac{P(\mathrm{X}_t = s|X_{t-1} = r, \mathbf{Z}_t)}{P(\mathrm{X}_t = 1|X_{t-1} = r, \mathbf{Z}_t)} = \gamma_{0s} + \gamma_{0rs} + \boldsymbol{\gamma}_s'\,\mathbf{Z}_t, \tag{6}$$

with $1 < s \leq S$, and $1 < r \leq S$ for the transitions probabilities. We take the first category as reference - setting to zero the related parameters. For the transition model, this means that parameters related to the elements in the first row and column of the transition matrix are set to zero. This produces $\underbrace{J \times S}_{\text{measurement model}} + \underbrace{(S-1) \times (K+1)}_{\text{initial state probabilities}} + \underbrace{(S-1) \times (1 + (S-1) + K)}_{\text{transition probabilities}}$ free parameters to be estimated.

The interpretation of category-specific effects is then in terms of difference from the first (reference) latent state.

Let $\{(\mathbf{Y}_{it}, \mathbf{Z}_{it})\}_n = \{(\mathbf{Y}_{1t}, \mathbf{Z}_{1t}), \ldots, (\mathbf{Y}_{nt}, \mathbf{Z}_{nt})\}$ be a sample of independent observations, observed for $t = 0, \ldots, T$ time occasions. Parameter estimation is normally carried out in one step by maximizing the log-likelihood function

$$L = \sum_{i=1}^{n} \log P(\mathbf{Y}_i = \mathbf{y}_i|\,\mathbf{Z}_i). \tag{7}$$

Iterative procedures, like the EM algorithm (Dempster et al., 1977) can be used to maximize Equation (7). However, when using the standard EM, the time and storage required for parameter estimation of LM models increases exponentially with the number of time points (Vermunt, Langeheine, & Böckenholt, 1999). For this reason, the forward–backward algorithm (Baum et al., 1970; Welch, 2003) is typically implemented: this is a special version of the standard EM 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 (Di Mari et al., 2016).

Stepwise estimation of latent Markov models

Bias-adjusted three-step method

Below we present the bias-adjusted three-step method of LM modeling as introduced by Di Mari et al.(2016). The first step model treats repeated measures from the same individual as independent. The pooled observations are modeled using a LC model as shown in the leftmost side of Figure 2 and defined as follows:

$$P(\mathbf{Y} = \mathbf{y}) = \sum_{s=1}^{S} P(\mathrm{X} = s) P(\mathbf{Y} = \mathbf{y} | \mathrm{X} = s), \tag{8}$$

where

$$P(\mathbf{Y} = \mathbf{y} | \mathrm{X} = s) = \prod_{j=1}^{J} P(\mathrm{Y}_j = \mathrm{y}_j | \mathrm{X} = s) \tag{9}$$

under local independence of responses given the latent process.

$P(X = s)$, the class proportions, and $P(\mathrm{Y}_j = \mathrm{y}_j | X = s)$, the conditional response probabilities, are parametrized with standard logistic regressions with only intercept terms, yielding to $J \times S + S - 1$ free parameters to be estimated. Estimation is done by maximizing the following log-likelihood function

$$L_{\text{STEP1}} = \sum_{i=1}^{n} \sum_{t=0}^{T} \log P(\mathbf{Y}_{it} = \mathbf{y}_{it}). \tag{10}$$

In the second step, individuals are assigned to latent classes based on their posterior probabilities as shown in the middle part of Figure 2. Posterior probabilities, or posterior class probabilities, are obtained as

$$P(\mathrm{X}_t = s_t | \, \mathbf{Y}_t = \mathbf{y}_t) = \frac{P(\mathrm{X}_t = s_t) P(\mathbf{Y}_t = \mathbf{y}_t | \mathrm{X}_t = s_t)}{P(\mathbf{Y}_t = \mathbf{y}_t)}, \tag{11}$$

by applying the Bayes rule. The most common classification rules, based on Equation (11), are modal assignment, which yields a hard partitioning of the data, and proportional assignment, which yields a soft (or crisp) partitioning. We will focus on modal assignment only[1]. Denoting the predicted class by W_t, which only depends on $\mathbf{Y}_t$, modal assignment estimates W_t allocating a weight

[1] In principle, both modal and proportional assignment rules can be used. Yet, proportional assignment may become infeasible with larger number of time points, as nonzero weights are allocated to all possible latent state patterns of each observation.

$\mathrm{w}_{yts} = P(\mathrm{W}_t = s_t | \mathbf{Y}_t = \mathbf{y}_t) = 1$ if $P(\mathrm{X}_t = s_t | \mathbf{Y}_t = \mathbf{y}_t)$ is the largest, and a zero weight otherwise.

The amount of classification error can be evaluated as the probability of the estimated class value given the true one, which, by aggregating over the sample patterns, can be obtained (Vermunt, 2010) as

$$P(\mathrm{W}_t = r_t | \mathrm{X}_t = s_t) = \frac{\frac{1}{n(T+1)} \sum_{i=1}^{n} \sum_{t=0}^{T} P(\mathrm{X}_t = s_t | \mathbf{Y}_t = \mathbf{y}_t) w_{its}}{P(\mathrm{X}_t = s_t)}. \tag{12}$$

In the third step (the rightmost part of Figure 2), class assignments are used to estimate the structural component of the model in Equation (1), while correcting for the classification error computed in step 2. Di Mari et al. (2016) showed that the ML correction of Vermunt (2010) for three-step estimation of LC models can be adapted for three-step estimation of LM models. That is, 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 $P(\mathrm{W}_t | \mathrm{X}_t)$ as known:

$$P(\mathbf{W} | \mathbf{Z}) = \sum_{X_0, \ldots, \mathrm{X}_t} P(\mathrm{X}_0 | \mathbf{Z}_0) \prod_{t=1}^{T} P(\mathrm{X}_t | \mathrm{X}_{t-1}, \mathbf{Z}_t) \prod_{t=0}^{T} P(\mathrm{W}_t | \mathrm{X}_t), \tag{13}$$

with $(S-1) \times (K+1) + (S-1) \times (1 + (S-1) + K)$ free parameters to be estimated.

Hence, the estimation of the parameters of the structural component of (1) is done by maximizing the following log-likelihood function:

$$L_{\text{STEP3}} = \sum_{i=1}^{n} \log \sum_{s_0, \ldots, s_T} P(\mathrm{X}_0 = s_0 | \mathbf{Z}_{i0}) \prod_{t=1}^{T} P(\mathrm{X}_t = s_t | \mathrm{X}_{t-1} = s_{t-1}, \mathbf{Z}_{it}) \prod_{t=0}^{T} P(\mathrm{W}_{it} = r_{it} | \mathrm{X}_t = s_t). \tag{14}$$

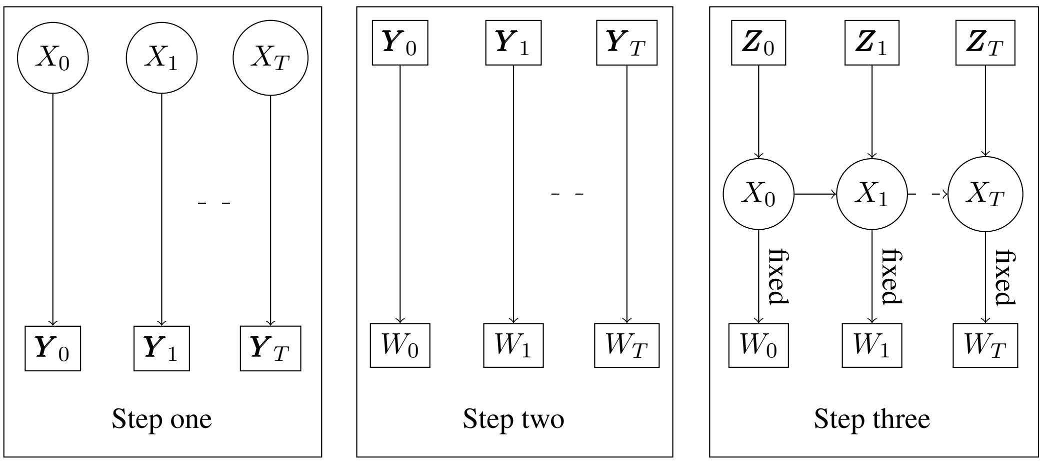

Figure 2. The steps of the three-step LM model

Two-step method with direct effects

An alternative estimator initially proposed by Bartolucci et al. (2014) is the two-step method. The first step model is a standard LC model estimated on the pooled data as described in Subsection “Bias-adjusted three-step method" and shown in the left hand side of Figure 3. Let us denote as $\boldsymbol{\theta}_1$ the conditional response probabilities, defined in Equation (9), which are estimated in the first step model. Bartolucci et al. (2015) showed that $\boldsymbol{\theta}_1$ can be used to consistently estimate the measurement model of Equation (2).

The second step consists in maximizing the following log-likelihood function

$$L(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1 = \hat{\boldsymbol{\theta}}_1) = \sum_{i=1}^{n} \log P(\mathbf{Y}_i = \mathbf{y}_i|\,\mathbf{Z}_i), \tag{15}$$

where we denote the $(S-1) \times (K+1) + (S-1) \times (1 + (S-1) + K)$ parameters of the initial state probabilities and transition probabilities as $\boldsymbol{\theta}_2$, and

$$P(\mathbf{Y}_i = \mathbf{y}_i|\,\mathbf{Z}_i) = \underbrace{\sum_{s_0=1}^{S} \sum_{s_1=1}^{S} \cdots \sum_{s_T=1}^{S} P(\mathrm{X}_0 = s_0|\,\mathbf{Z}_0) \prod_{t=1}^{T} P(\mathrm{X}_t = s_t|\mathrm{X}_{t-1} = s_{t-1}, \mathbf{Z}_{it})}_{\text{free}} \underbrace{\prod_{t=0}^{T} P(\mathbf{Y}_{it} = \mathbf{y}_{it}|\mathrm{X}_t = s_t)}_{\text{fixed}}. \tag{16}$$

The log-likelihood of Equation (15) is maximized with respect to $\boldsymbol{\theta}_2$, whereby the $J \times S$ parameters of the measurement model are fixed at their first step values $\hat{\boldsymbol{\theta}}_1$ (see the right hand side of Figure 3), and all direct effects of the covariates on the indicators are set to zero.

While Bartolucci et al. (2014) proposed a version of the model without the DEs, we hereby propose a natural extension of their approach that enables modeling DEs. To test whether fitting a model without any DEs causes any misfit, a local fit measure like the Longitudinal BiVariate Residuals (L-BVR) statistic (J. K. Vermunt & Magidson, 2016) can be used. If any direct effect of the covariates is picked up by the L-BVR, the indicators showing residual associations with $\mathbf{Z}$ are freed and the first step model is re-estimated by modeling the direct effect as follows:

$$P(\mathbf{Y} = \mathbf{y}) = \sum_{s=1}^{S} P(\mathrm{X} = s) P(\mathbf{Y} = \mathbf{y}|\mathrm{X} = s, \mathbf{Z}). \tag{17}$$

The conditional response probabilities of indicators showing no residual association are simply modeled by using $P(\mathbf{Y} = \mathbf{y}|\mathrm{X} = s)$ instead of $P(\mathbf{Y} = \mathbf{y}|\mathrm{X} = s, \mathbf{Z})$. The estimated parameters are then used as fixed-value parameters in the second step, where the log-likelihood function of Equation (15) is maximized again with respect to $\boldsymbol{\theta}_2$. The estimation is broken down into peaces, making complex LM modeling with covariates manageable, whereas one-step estimation would be cumbersome, especially with many time points, indicators and/or covariates.

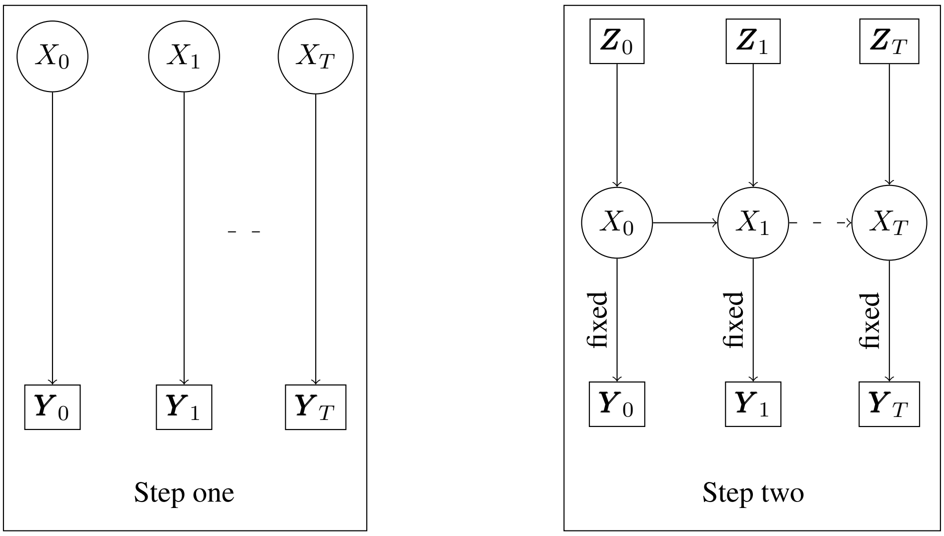

Figure 3. The steps of the two-step LM model

The augmented bias-adjusted three-step approach

Unbiased three-step modeling in the presence of direct effects can be done if all residual associations are modeled correctly. Because the use of residual statistics to identify direct effects is not possible using the three-step approach, we propose an alternative somewhat overparametrized approach that could potentially obtain similar results to correctly modeling the (known) direct effects. Our approach is motivated by the work of Asparouhov & Muthén (2014), who proposed an extension to the three-step LC model that allows modeling direct effects by allowing for them in step one.

Let us introduce an additional latent variable Ψ, with state space of size P and we let p be one of the possible states. In addition, let us define D as the set of the J^*, with $J^* \leq J$, indicators for which we might presume direct effects of the covariates. Instead of the model in Equation (8), we consider modeling the pooled sample according to the following augmented first step model

$$P(\mathbf{Y} = \mathbf{y}) = \sum_{s=1}^{S} P(\mathrm{X} = s) P(\mathbf{Y} = \mathbf{y} | \mathrm{X} = s, \mathbf{Z}, \Psi = p), \tag{18}$$

where

$$P(\mathbf{Y} = \mathbf{y} | \mathrm{X} = s, \mathbf{Z}, \Psi = p) = \prod_{j \notin D} P(\mathrm{Y}_j = \mathrm{y}_j | \mathrm{X} = s) \times \prod_{j \in D} P(\mathrm{Y}_j = \mathrm{y}_j | \mathrm{X} = s, \mathbf{Z}, \Psi = p). \tag{19}$$

The parametrization we consider for the response probabilities is as follows

$$\log\left(\frac{\pi_{sj}}{1 - \pi_{sj}}\right) = \theta_{js} + \mathbb{1}\{j \in D\}(\psi_{jp} + \boldsymbol{\alpha}_j' \, \mathbf{Z}), \tag{20}$$

for $1 \leq s \leq S$ and $1 \leq p \leq P$, with $J \times S + J^* \times P + J^* \times K$ free parameters. In most instances researchers have an initial knowledge of which indicators are prone to have bias. In the augmented first step model of Equation (18), the covariates assumed to have a direct effect are loaded on the indicators (see Figure 4) based on a - even rough - knowledge on where the direct effects can be located. Intuitively, the additional latent variable Ψ is included in order to model the noise (D. Oberski, 2016) that can be caused by misspecifying the direct effects. We choose to model such noise non-parametrically by using a discrete latent variable, similarly to what is done in the LC latent factor models of J. Vermunt & Magidson (2004).

Once the step one model is estimated, posterior probabilities are computed using the formula in Equation (11), and modal class assignments are then used - correcting for the classification error - to estimate the structural component of the model in Equation (1).

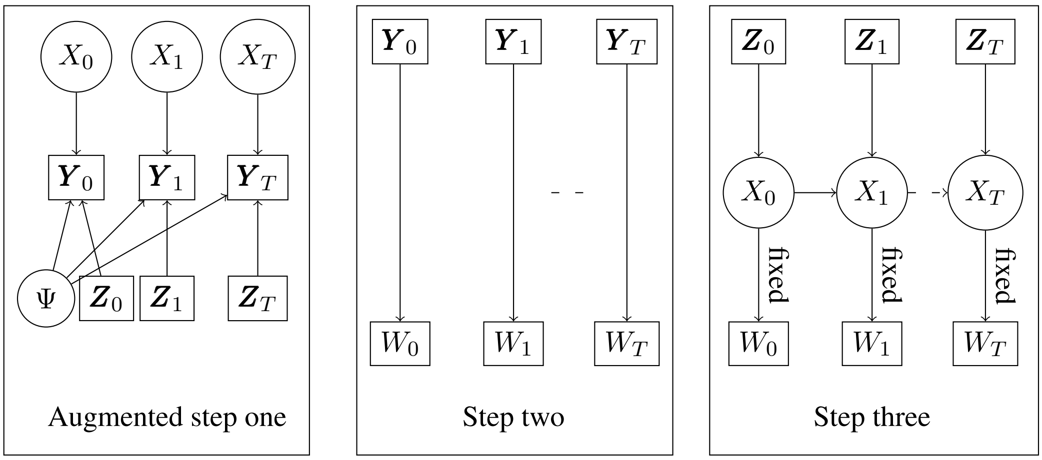

Figure 4. The steps of the augmented three-step LM model

Simulation study

Design

The objective of this simulation study is to evaluate the quality of the different modeling approaches for latent Markov models with covariates in the presence of one or more direct effects (DE) of the covariates on the indicators. Next to evaluating the commonly used approaches (one-step, two-step and three-step), we compare their performance to the hereby proposed modified stepwise estimators, that model the direct effects. Our expectation is that the approaches that model the direct effects (either explicitly or by modeling them as noise) will do better then the approaches that ignore these effects, especially when they are strong in magnitude, and affect many indicators. We restrict ourselves to the situation where we model known direct effects, ignoring the problem of correctly selecting them. Indeed, our main interest is to investigate the percentage bias caused by the presence of direct effects, and whether this bias can be corrected if such direct effects are correctly modeled.

Each method will be assessed in terms of parameter estimates, percentage bias, relative efficiency (averaged standard error over averaged standard deviation), coverage rates and relative loss of precision in the estimate with respect to the correctly specified one-step (FIMLcorr) model. The methods included in the comparison are 1) one-step without DE (FIMLuncor), 2) correctly specified one-step (FIMLcorr), 3) two-step without DE (2STEP uncor) and 4) two-step with DE (2STEP corr), 5) bias-adjusted three-step with unmodeled DE (3STEP), 6) augmented bias-adjusted three-step (3STEPaug), where the second LV - discrete, with 6 latent states[2] - and the covariate effects are loaded only on the three items which might have residual association with the covariates.

Class separation - strongly related to classification error - and sample size are known to affect stepwise modeling. In this simulation study, we manipulate separation only through the strength of the relationship between items and latent classes - conditional response probabilities - although other options are possible - e.g. number of items, the number of item categories, and the class sizes.

We generate the data from a three-class latent Markov model with six dichotomous responses (low/high) and two numeric time-fixed covariates, where one (Z_1) is included in both the initial state and the transition model, while the other (Z_2) is only included in the transition model. Z_1 is binary with scores -0.5, 0.5, while Z_2 has five categories, scored -2, -1, 0, 1, 2, and treated as continuous. Covariates Z_1 and Z_2 enter the structural models through the multinomial logit parametrization shown in Equations (5) and (6), with the first class as reference category. Parameter values are chosen such that initial state probabilities are about the same, and the probability of switching is about 0.2 - a typical setup in LM modeling, where stability tends to be larger than change. The intercepts of the initial state model were both set to 0, while $\beta_{2Z_1} = \beta_{3Z_1} = -0.5$. In 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 to -1 for Z_1 in both classes, and to 0.25 for Z_2 in both classes.

Z_1 has also a direct effect on the indicators. We consider two direct effect conditions: one, where the effect is stronger in magnitude (0.5), and another one where the effect is milder (0.25). For each of the two magnitude conditions, we consider three sub-scenarios, in which Z_1 is loaded on one, two and three indicators respectively. For instance, for the strong direct effect condition, the parametrization is as in Equation (4), with (sub-scenario 1) a DE on the first item $\alpha_1 = 0.5$;

[2]A larger number of states was also tried, but in all cases we observed that a number approximately twice as large as the number of states in the latent variable of interest was enough to approximate the distribution of the noise.

(sub-scenario 2) DE's on the first and the last items, with $\alpha_1 = \alpha_6 = 0.5$; (sub-scenario 3) DE's on the first, the third and the last items, with $\alpha_1 = \alpha_6 = 0.5$, and $\alpha_3 = -0.5$. All other α's are set to zero. The scenarios, for a given magnitude (0.25 or 0.5) are displayed in Table 1.

Classes' profiles are as follows: 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 (Table 1). The response probabilities for the most likely responses were set to 0.8 and 0.9, corresponding to moderate - entropy-based R^2 of about 0.7 - and high class separation - entropy-based R^2 of about 0.9 - in the first step model. The sample sizes used were 100, 500 and 1,000, and each sample unit is observed for 5 measurement occasions.

For all 36 simulation conditions, obtained by combining the specified class separation levels, sample sizes, and the 3×2 direct effect scenarios, we considered 500 replications and estimated the model using all six methods. Data generation and parameter estimation were carried out using Latent GOLD version 5.0 (J. K. Vermunt & Magidson, 2013), and results were analyzed in R. Results are presented in Subsection "Results".

Table 1
Item-class probabilities in the simulation, with associated direct effects and sign in parentheses.

Item	Class 1	Class 2	Class 3	1 D.E.	2 D.E.	3 D.E.
1	low	high	high	✓(+)	✓(+)	✓(+)
2	low	high	high	×	×	×
3	low	high	low	×	×	✓(-)
4	high	low	low	×	×	×
5	low	low	high	×	×	×
6	high	low	high	×	✓(+)	✓(+)

Results

Tables 2 and 3 present percentage bias, average relative efficiency (SE/SD), coverage of 95% CI's (CVG) and relative loss of parameter estimates with respect to FIML corr for all six methods, averaged over covariate effects, sample size and separation conditions for the three DE scenarios, respectively for DE's magnitude of 0.25 and 0.50. As can be seen, the bias-adjusted three-step (3STEP) does systematically better than the one-step (FIML uncor) approach, and similarly to the two-step (2STEP uncor) with unmodeled DE approach in terms of bias. However all uncorrected estimators have a percentage bias up to as high as 6% in the conditions with two and three direct effects, in the strong magnitude case - 4 % in the mild magnitude case. At the same time the two-step approach with modeled DE (2STEP corr) shows the same performance as the correctly specified one-step approach (FIML corr) in all DE scenarios, having a % bias of 2 / 2.5 percent in all conditions. The augmented three-step (3STEPaug) represents a good alternative - in terms of bias, SE/SD, coverage and loss of accuracy with respect to FIML corr - to the two-step approach with modeled DE, with difference between the two approaches vanishing as the number of DE's increases. With all estimators, the SE/SD is close enough to 1, showing that the SE estimators approximate the sampling variance well. Although, in some conditions with DE's magnitude of

0.25, there is a slight underestimation of the standard errors (SE/SD of about 0.98), the coverage is close to the 95% for all estimators.

Table 2

Simulation Results - Covariate percentage bias (%Bias), SE/SD, coverage rates (CVG), and relative loss - RelLoss - of parameter estimates with respect to the correctly specified model (FIMLcorr), averaged over all covariate effects and simulation conditions. Magnitude of direct effect = mild *(0.25)*

	%Bias	SE/SD	CVG	RelLoss
	1 direct effect			
FIMLcorr	0.019	0.978	0.945	-
FIMLuncor	0.025	0.978	0.946	1.002
2STEP uncor	0.022	0.980	0.946	0.997
2STEPcorr	0.022	0.990	0.948	1.001
3STEP	0.022	0.979	0.946	1.000
3STEPaug	0.028	0.982	0.946	0.984
	2 direct effects			
FIMLcorr	0.023	0.981	0.948	-
FIMLuncor	0.044	0.981	0.946	1.021
2STEP uncor	0.040	0.983	0.946	1.016
2STEPcorr	0.019	0.984	0.948	0.994
3STEP	0.042	0.988	0.948	1.018
3STEPaug	0.024	0.987	0.949	0.992
	3 direct effects			
FIMLcorr	0.022	0.996	0.953	-
FIMLuncor	0.039	0.996	0.952	1.020
2STEP uncor	0.035	0.998	0.950	1.014
2STEPcorr	0.019	0.999	0.952	0.993
3STEP	0.036	0.996	0.952	1.016
3STEPaug	0.027	0.997	0.951	0.993

Next we zoom in on the results separately for each combination of sample size, separation and number of direct effects, for both magnitude levels. We report the % bias, SE/SD and coverage for one of the parameter estimates measuring the covariate effect on the initial state ($\beta_{2Z_1} = -0.50$). We do so because the amount of bias we observe in the covariate effects on the initial states is larger than on the transition parameters, so the least optimistic results are reported.

Tables 4 and 7 - magnitudes of 0.25 and 0.50 respectively - show that in the one direct effect condition the approaches that ignore the direct effects are doing well, having small percentage bias in all situations. These results are in line with the findings of Asparouhov & Muthén (2014), who also found that in the one direct effect condition ignoring the DE does not cause large bias, contrary to the conditions with more DEs. Similarly to the results reported in Tables 2 and 3, the coverage and SE/SD shows values close to nominal levels in all situations.

Table 3

Simulation Results - Covariate percentage bias (%Bias), SE/SD, coverage rates (CVG), and relative loss - RelLoss - of parameter estimates with respect to the correctly specified model (FIMLcorr), averaged over all covariate effects and simulation conditions. Magnitude of direct effect = strong *(0.50)*

	%Bias	SE/SD	CVG	RelLoss
	1 direct effect			
FIMLcorr	0.025	0.998	0.953	-
FIMLuncor	0.036	0.997	0.952	1.004
2STEP uncor	0.033	1.000	0.952	0.998
2STEPcorr	0.019	0.991	0.950	0.996
3STEP	0.035	1.005	0.955	0.999
3STEPaug	0.034	0.988	0.948	0.978
	2 direct effects			
FIMLcorr	0.021	0.998	0.953	-
FIMLuncor	0.057	0.997	0.949	1.039
2STEP uncor	0.051	0.999	0.948	1.037
2STEPcorr	0.016	1.001	0.953	0.993
3STEP	0.053	1.004	0.950	1.035
3STEPaug	0.025	1.005	0.951	1.001
	3 direct effects			
FIMLcorr	0.026	0.994	0.950	-
FIMLuncor	0.060	0.992	0.946	1.041
2STEP uncor	0.054	0.994	0.947	1.037
2STEPcorr	0.021	0.996	0.950	0.993
3STEP	0.057	0.994	0.949	1.039
3STEPaug	0.032	0.997	0.951	1.009

Some interesting differences stand out in the two and three direct effect conditions, for both magnitude levels. For the stronger magnitude, we observe (Tables 8 and 9) the following patterns: the uncorrected approaches have a percentage bias up to 17.5% for the FIML approach and 14.5% for the stepwise approaches in the smallest sample size and moderate separation condition for the two direct effect condition, and up to 24% for FIML, and 22% for the stepwise estimators in the three direct effect condition. Overall, among the approaches that do not model the direct effects, the FIMLuncorr has the most bias, followed by the 2stepuncor, with the most robust estimator being the 3STEP. With modeling the direct effects, all estimators have a small amount of bias, and there is no clear ordering of their performance with regard to bias. Furthermore, in all conditions, the SE/SD and coverage rate are close to the nominal level.

For the milder magnitude scenarios, with two and three direct effects (Tables 5 and 6), values for percentage bias are relatively smaller than for stronger magnitude, although we observe similar patterns. Interestingly, all estimators, in some conditions of small sample size and/or moderate class

Table 4

Simulation Results - Covariate percentage bias (%Bias), SE/SD, coverage rates (CVG) - of $\beta_2 Z_1$ =(-0.50) separatly for all combinations of sample size and separation for the one direct effect condition, magnitude 0.25.

	Moderate separation								
	n=100			n=500			n=1000		
	%Bias	SE/SD	CVG	%Bias	SE/SD	CVG	%Bias	SE/SD	CVG
FIMLcorr	0.008	0.970	0.948	0.022	0.981	0.938	0.011	0.973	0.948
FIMLuncorr	0.029	0.967	0.948	0.010	0.983	0.942	0.021	0.974	0.950
2STEPuncor	0.042	0.978	0.944	0.014	0.985	0.942	0.023	0.976	0.948
2STEPcorr	0.037	0.971	0.954	0.020	1.015	0.948	0.012	0.987	0.950
3STEP	0.007	0.993	0.958	0.003	0.976	0.944	0.025	0.982	0.950
3STEPaug	0.023	1.006	0.958	0.043	0.969	0.940	0.081	0.985	0.942
	Large separation								
	n=100			n=500			n=1000		
	%Bias	SE/SD	CVG	%Bias	SE/SD	CVG	%Bias	SE/SD	CVG
FIMLcorr	0.053	0.961	0.940	0.025	1.060	0.964	0.003	0.963	0.936
FIMLuncorr	0.070	0.961	0.936	0.042	1.060	0.964	0.020	0.962	0.942
2STEPuncor	0.076	0.971	0.946	0.043	1.064	0.964	0.020	0.963	0.942
2STEPcorr	0.094	1.032	0.966	0.019	0.989	0.956	0.013	1.017	0.950
3STEP	0.065	0.966	0.944	0.042	1.068	0.968	0.016	0.966	0.938
3STEPaug	0.065	0.966	0.940	0.052	1.067	0.960	0.032	0.962	0.934

separation, show some degree of underestimation of the SE's - the smallest SE/SD ratio reported being of about 0.93 in the 1000 observations and moderate separation condition. It seems like, in the milder magnitude condition, if more than one direct effect is present, this tends to produce larger variability in the estimates of the parameters of interest. Nonetheless, the underestimation in the SE's is negligible as coverage rates are of $\approx$ 0.95 in all cases.

Table 5

Simulation Results - Covariate percentage bias (%Bias), SE/SD, coverage rates (CVG) - of β_2(-0.50) separatly for all combinations of sample size and separation for the two direct effect condition, magnitude 0.25.

	Moderate separation								
	n=100			n=500			n=1000		
	%Bias	SE/SD	CVG	%Bias	SE/SD	CVG	%Bias	SE/SD	CVG
FIMLcorr	0.138	1.026	0.960	0.004	0.963	0.936	0.018	0.933	0.940
FIMLuncorr	0.183	1.022	0.958	0.050	0.966	0.944	0.063	0.938	0.936
2STEPuncor	0.172	1.037	0.964	0.046	0.969	0.944	0.061	0.939	0.938
2STEPcorr	0.114	1.048	0.968	0.011	0.969	0.936	0.013	0.937	0.942
3STEP	0.186	1.048	0.956	0.039	0.988	0.940	0.060	0.938	0.942
3STEPaug	0.122	1.061	0.966	0.023	0.985	0.942	0.028	0.940	0.942
	Large separation								
	n=100			n=500			n=1000		
	%Bias	SE/SD	CVG	%Bias	SE/SD	CVG	%Bias	SE/SD	CVG
FIMLcorr	0.001	0.952	0.942	0.019	0.980	0.938	0.015	1.048	0.962
FIMLuncorr	0.022	0.951	0.948	0.039	0.982	0.942	0.036	1.051	0.956
2STEPuncor	0.019	0.962	0.946	0.038	0.984	0.940	0.035	1.053	0.956
2STEPcorr	0.008	0.965	0.946	0.017	0.982	0.940	0.013	1.050	0.964
3STEP	0.024	0.963	0.946	0.038	0.976	0.940	0.033	1.058	0.966
3STEPaug	0.026	0.965	0.952	0.047	0.975	0.936	0.033	1.065	0.968

Table 6

Simulation Results - Covariate percentage bias (%Bias), SE/SD, coverage rates (CVG) - of β_2(-0.50) separatly for all combinations of sample size and separation for the three direct effect condition, magnitude 0.25.

	Moderate separation								
	n=100			n=500			n=1000		
	%Bias	SE/SD	CVG	%Bias	SE/SD	CVG	%Bias	SE/SD	CVG
FIMLcorr	0.003	1.046	0.964	0.007	1.040	0.962	0.014	1.040	0.962
FIMLuncorr	0.095	1.043	0.960	0.087	1.043	0.960	0.069	1.040	0.962
2STEPuncor	0.077	1.057	0.970	0.083	1.046	0.956	0.066	1.043	0.962
2STEPcorr	0.025	1.066	0.966	0.002	1.043	0.964	0.018	1.045	0.964
3STEP	0.056	1.050	0.968	0.087	1.033	0.944	0.062	1.040	0.958
3STEPaug	0.036	1.049	0.962	0.064	1.030	0.950	0.034	1.015	0.948
	Large separation								
	n=100			n=500			n=1000		
	%Bias	SE/SD	CVG	%Bias	SE/SD	CVG	%Bias	SE/SD	CVG
FIMLcorr	0.103	0.989	0.958	0.005	0.999	0.950	0.014	0.981	0.958
FIMLuncorr	0.143	0.988	0.958	0.031	0.994	0.938	0.026	0.980	0.952
2STEPuncor	0.133	0.998	0.958	0.030	0.995	0.938	0.025	0.982	0.954
2STEPcorr	0.090	1.001	0.962	0.008	1.001	0.952	0.015	0.984	0.958
3STEP	0.127	1.008	0.962	0.030	0.982	0.946	0.020	0.994	0.946
3STEPaug	0.113	1.009	0.956	0.026	0.982	0.950	0.012	0.986	0.950

Table 7

Simulation Results - Covariate percentage bias (%Bias), SE/SD, coverage rates (CVG) - of $\beta_2 Z_1$ =(-0.50) separatly for all combinations of sample size and separation for the one direct effect condition, magnitude 0.50.

	Moderate separation								
	n=100			n=500			n=1000		
	%Bias	SE/SD	CVG	%Bias	SE/SD	CVG	%Bias	SE/SD	CVG
FIMLcorr	0.054	0.966	0.960	0.017	1.019	0.958	0.004	1.063	0.964
FIMLuncorr	0.008	0.964	0.960	0.079	1.016	0.954	0.057	1.065	0.962
2STEPuncor	0.027	0.975	0.958	0.083	1.022	0.954	0.059	1.067	0.962
2STEPcorr	0.085	0.969	0.948	0.037	1.089	0.978	0.017	0.982	0.942
3STEP	0.052	0.986	0.966	0.079	1.006	0.958	0.064	1.089	0.970
3STEPaug	0.124	1.016	0.962	0.099	0.995	0.954	0.122	0.985	0.926
	Large separation								
	n=100			n=500			n=1000		
	%Bias	SE/SD	CVG	%Bias	SE/SD	CVG	%Bias	SE/SD	CVG
FIMLcorr	0.030	1.035	0.974	0.029	0.973	0.942	0.020	1.044	0.950
FIMLuncorr	0.001	1.035	0.972	0.003	0.974	0.942	0.052	1.043	0.952
2STEPuncor	0.009	1.046	0.976	0.005	0.974	0.942	0.053	1.044	0.950
2STEPcorr	0.026	1.061	0.974	0.006	0.998	0.952	0.003	1.024	0.960
3STEP	0.009	1.049	0.978	0.002	0.978	0.942	0.050	1.038	0.950
3STEPaug	0.018	1.036	0.956	0.052	0.969	0.932	0.027	0.988	0.940

Table 8

Simulation Results - Covariate percentage bias (%Bias), SE/SD, coverage rates (CVG) - of β_2(-0.50) separatly for all combinations of sample size and separation for the two direct effect condition, magnitude 0.50.

	Moderate separation								
	n=100			n=500			n=1000		
	%Bias	SE/SD	CVG	%Bias	SE/SD	CVG	%Bias	SE/SD	CVG
FIMLcorr	0.070	0.928	0.944	0.008	1.019	0.962	0.019	1.013	0.956
FIMLuncorr	0.175	0.934	0.942	0.091	1.013	0.960	0.081	1.015	0.950
2STEPuncor	0.145	0.956	0.950	0.084	1.014	0.958	0.075	1.012	0.946
2STEPcorr	0.032	0.953	0.950	0.016	1.021	0.964	0.024	1.011	0.956
3STEP	0.143	0.974	0.952	0.087	1.014	0.958	0.070	0.984	0.942
3STEPaug	0.091	0.991	0.950	0.068	1.022	0.958	0.031	0.965	0.942
	Large separation								
	n=100			n=500			n=1000		
	%Bias	SE/SD	CVG	%Bias	SE/SD	CVG	%Bias	SE/SD	CVG
FIMLcorr	0.032	0.970	0.954	0.033	1.033	0.954	0.027	1.036	0.950
FIMLuncorr	0.013	0.972	0.954	0.076	1.031	0.944	0.071	1.040	0.950
2STEPuncor	0.005	0.989	0.960	0.075	1.032	0.944	0.069	1.042	0.952
2STEPcorr	0.043	0.987	0.960	0.032	1.035	0.956	0.025	1.039	0.95
3STEP	0.004	1.003	0.964	0.073	1.034	0.938	0.069	1.043	0.946
3STEPaug	0.012	0.999	0.962	0.076	1.025	0.930	0.068	1.047	0.950

Table 9

Simulation Results - Covariate percentage bias (%Bias), SE/SD, coverage rates (CVG) - of β_2(-0.50) separatly for all combinations of sample size and separation for the three direct effect condition, magnitude 0.50.

	Moderate separation								
	n=100			n=500			n=1000		
	%Bias	SE/SD	CVG	%Bias	SE/SD	CVG	%Bias	SE/SD	CVG
FIMLcorr	0.054	1.029	0.946	0.032	1.021	0.954	0.025	1.043	0.948
FIMLuncorr	0.242	1.023	0.952	0.142	1.021	0.948	0.147	1.028	0.930
2STEPuncor	0.225	1.037	0.956	0.136	1.028	0.946	0.140	1.030	0.928
2STEPcorr	0.029	1.044	0.960	0.037	1.029	0.954	0.029	1.043	0.950
3STEP	0.214	1.038	0.964	0.140	1.047	0.954	0.148	1.013	0.930
3STEPaug	0.025	1.074	0.968	0.095	1.042	0.956	0.096	0.998	0.946
	Large separation								
	n=100			n=500			n=1000		
	%Bias	SE/SD	CVG	%Bias	SE/SD	CVG	%Bias	SE/SD	CVG
FIMLcorr	0.026	0.981	0.960	0.012	0.946	0.936	0.002	1.016	0.964
FIMLuncorr	0.054	0.981	0.960	0.068	0.945	0.936	0.083	1.017	0.946
2STEPuncor	0.047	0.995	0.964	0.066	0.949	0.938	0.081	1.017	0.946
2STEPcorr	0.033	0.995	0.958	0.014	0.951	0.938	0.001	1.017	0.964
3STEP	0.042	1.011	0.962	0.062	0.957	0.944	0.076	1.020	0.946
3STEPaug	0.018	1.008	0.952	0.059	0.963	0.946	0.067	1.029	0.948

A latent Markov model of tolerance types

Our choice of real data example is motivated by an example used by McCutcheon (McCutcheon, 1985) who showed how age and education groups differ with regard to their tolerance toward minority groups on data from the 1976 and 1977 General Social Survey (GSS) using a naive three-step approach.

We used data from the panel version of the same GSS survey including measurements from three years: 2010, 2012 and 2014, with a sample size of 2044, 1551 and 1304 respectively. The data is openly available[3]. We selected six items that measure whether the respondents would allow members of different out-groups to speak in a public space. The items are formulated as follows: "Suppose this ... wanted to make a speech in your community. Should he be allowed to speak?" with response options "(**Yes**/No)". The six items refer to: communists, atheists, militarists, homosexuals, racists and Muslims. These items are used to measure an aspect of tolerance toward out-groups. It is interesting to note that, in the 1976 /77 version of the questionnaire, the item measuring tolerance toward Muslims was not yet included.

The covariates were re-coded following the approach of McCutcheon. Education was coded into three categories: less than twelve degrees (1), completed highschool (2) and higher educated (3). Cohort was coded into four categories: old (62 years and older, group 4), middle age (43-61 years old, group 3), young adults (25-42 years old, group 2) and young (18-24 years old, group 1).

We applied all six approaches to analyze the data. The results are discussed below.

Using both stepwise approaches in step one, following the recommendations of Di Mari et al. (2016), we run a latent class model on the pooled data from the three years on the six items representing the measurement model. Using this approach we assume a time invariant measurement model - an approach commonly used in stepwise modeling, known to be robust. In Table 10 the class sizes and class specific response probabilities of the best fitting three class model are presented (AIC=-4150.90, BIC= -24771.78, χ^2= 4192.30).

[3] https://www.icpsr.umich.edu/icpsrweb/ICPSR/series/28/studies/7398?archive=ICPSR&sort

Table 10
Class sizes and conditional probabilities to give all-tolerant answers given the latent class.

	"Intolerant"	"Tolerant"	"Middle"
Class size	0.19	0.51	0.30
Allow to speak...			
Atheists	0.17	0.99	0.84
Communists	0.06	0.99	0.56
Militarists	0.16	0.97	0.64
Racists	0.13	0.94	0.44
Homosexuals	0.52	0.99	0.92
Muslims	0.02	0.86	0.07

The first class, labeled ‘Intolerant’, has a low probability to give a positive answer on all items, while the second class (which is the largest class) has a high probability to give a positive answer on all items. The last class, labeled ‘Middle’, has a low probability to allow to speak for Muslims and racists, while being more tolerant toward the other groups.

Using the three-step approach, after establishing the step one model, posterior classifications are obtained in step two, and in step three the Markov model is built using the posterior classifications while correcting for the classification error. We used two implementations: the three-step approaches without correcting for possible direct effects, and the augmented three-step approach. As for the latter option, we decided to model possible direct effects by regressing the predictors and the second latent variable (with 6 classes) on 3 items that can have direct effects in the first step. Using this more complex step one model the posterior classifications were obtained in step two and used in the step three analysis. The Latent GOLD syntax for the more complex augmented three-step approach is available in the appendix together with the syntax of the two-step approach modeling the direct effect.

Using the two-step approach, after establishing the LC model, in step two we estimate the transition probabilities of the Markov model, and the effect of the covariates on the initial states and on the transition probabilities. After estimating this model we check for residual dependencies between the covariates and the six indicator variables using the L-BVR statistic (J. K. Vermunt & Magidson, 2016). Two high residuals are identified between ’homosexuals’ and cohort (L-BVR=25.50) and between ’racists’ and education (L-BVR=14.10). The presence of these residuals shows that, after controlling for the latent classification, there is still a strong association left between the covariates and these specific items that should be accounted for. In order to account for these direct effects, we refit the step one model including the direct effects, followed by the corrected second step.

Using the one-step approach a similar strategy is used as with the two-step approach. After fitting the FIML model ignoring all possible direct effects we checked the residual statistics. Using the L-BVR the same direct effects were identified as with using the two-step approach, namely between ’homosexuals’ and cohort (L-BVR=25.50) and between ’racists’ and education (L-BVR=14.10). After identifying the direct effects the model was refitted by modeling these effects explicitly.

It should be mentioned that we show the use of residual statistics mainly for illustrative purposes. We recommend researchers not to use these statistics blindly, but in conjunction with theory.

In Table 11 we report the model fit statistics for the one-step and two-step approaches with and without modeling the direct effects. Using both approaches we can see a significant decrease in the L^2 value, which is also reflected in a decrease in the AIC, but not consistently in the BIC that controls for model complexity. Given the significant change in the L^2, we proceed with the more complex model that includes the direct effects. In three-step modeling, such a model selection strategy cannot be used, as the step-one and step-three models are not nested.

The estimated transition probabilities from the step one model of remaining in the same class are .70 for class one and three (intolerant and middle), and .90 for class two (tolerant class). We found that 25% of the respondents moved from class one (intolerant) to the moderately tolerant class (class three). Furthermore, the second largest switch is the transition over time from the middle to the tolerant class (18% of the respondents).

In Table 12 we report the covariate effects on the initial states and transition probabilities using the different approaches. More educated people have a higher probability to be in the middle and the tolerant group than the intolerant in the initial state. Education has a similar positive effect also on the transition probabilities. The effect of cohort is similar. Younger people have a higher probability

Table 11
Model fit statistics with the one-step and two-step approach

	AIC	BIC	L^2	df difference	L^2 dif	p
1-step	3660.57	-7640.98	7680.57			
1-step direct	3597.64	-7692.66	7613.64	2	66.93	<.001
2-step	3692.32	-7710.44	7748.32			
2-step direct	3646.46	-7711.32	7685.46	2	62.76	<.001

to be in the tolerant and middle group as compared to the intolerant group at the initial state. Cohort has a significant effect on the transition probabilities with both approaches, namely younger people have a higher probability to be in the tolerant and middle classes as compared to intolerant over time. Furthermore, after controlling for class, there is a direct effect of cohort on education, namely older people are more tolerant toward homosexuals then predicted only by the restrictive model. Using all the approaches, the same overall conclusions are reached with regard to the direction of the effects.

Conclusion

In the current paper we addressed an issue which is genuinely overlooked in latent Markov modeling: the consequences of ignoring direct effects between indicators of the latent concept and covariates. While in literature on latent class analysis and regression mixtures it is well known that direct effects can bias the parameters of interest (Kim et al., 2016; Asparouhov & Muthén, 2014), there is no literature on this issue in latent Markov modeling. The reason can be that LM models are very complex, and estimating these models ignoring the direct effects is already problematic. We compared the performance of the three estimators used in LM modeling: the one-step approach (also known as full information ML), and the stepwise estimators, the two- and three-step approaches. Furthermore we proposed two improved stepwise approaches that can take into account the direct effects while keeping the main benefit of stepwise estimation.

We compared the traditional approaches, namely the one-step approaches and the stepwise approaches, with the two- and three-step approaches with the hereby proposed augmented first step, that model the direct effects instead of ignoring them. All methods were compared to the one-step approach including the direct effects, which, although being the statistically most efficient and unbiased estimator, is in practice hardly ever used because of the complexities of model estimation.

The results of the simulation study show that ignoring the direct effects will lead to bias in the estimated relationship between the LC variable and the covariate effects. Even more, the bias is more severe with increasing magnitude and number of direct effects. Both stepwise corrected approaches (modeling the direct effects) do better than the uncorrected approaches in all conditions, and they have estimates similar to the ones obtained with the correctly specified one-step model. The two-step approach follows the one-step corrected closer than the three-step in the one and two direct effect conditions, however the differences are less then 1 % averaged over all conditions. Furthermore it is also relevant to see that, among the incorrectly specified models, the three-step approach has the least bias, so is the most robust to misspecifications. We also found that, with at least two direct effects with magnitude of 0.25 and under conditions of moderate class separation and/or small sample size, all estimators show larger sampling variability and SEs tend to be slightly

Table 12

The covariate effects on the initial states and transition probabilities obtained using the different approaches with class one as reference category

	FIMLcorr		2stepcorr		3STEPaug		FIMLuncorr		2stepuncorr		3STEP	
educ_state1	1.46	(0.13)	1.41	(0.12)	1.20	(0.13)	1.41	(0.12)	1.32	(0.12)	1.27	(0.13)
educ_state2	0.66	(0.12)	0.47	(0.12)	0.20	(0.14)	0.62	(0.11)	0.43	(0.12)	0.38	(0.13)
cohort_state1	-0.19	(0.09)	-0.19	(0.10)	-0.33	(0.11)	-0.23	(0.09)	-0.33	(0.10)	-0.32	(0.11)
cohort_state2	-0.10	(0.10)	-0.12	(0.11)	-0.33	(0.14)	-0.17	(0.10)	-0.25	(0.12)	-0.29	(0.13)
educ_transition	0.80	(0.17)	0.79	(0.16)	0.71	(0.19)	0.77	(0.17)	0.83	(0.17)	0.83	(0.18)
educ_transition	0.42	(0.12)	0.48	(0.14)	0.64	(0.21)	0.42	(0.13)	0.60	(0.15)	0.71	(0.21)
cohort_transition	-0.31	(0.13)	-0.39	(0.13)	-0.56	(0.16)	-0.32	(0.13)	-0.39	(0.15)	-0.50	(0.16)
cohort_transition	-0.24	(0.11)	-0.26	(0.12)	-0.40	(0.16)	-0.27	(0.11)	-0.40	(0.14)	-0.47	(0.15)

underestimated. Although a SEs correction is available (Bakk et al., 2014; Di Mari et al., 2016), given the already complex modeling involved and the almost perfect coverage rates found in all simulation conditions, we suggest not using it.

The approach we propose for direct effect modeling in a latent Markov context are mostly data-driven, in that they require refitting the model on the data at least twice and there is no theoretical basis to specify the direct effects. Although this cannot be done carelessly, such an approach is quite common in LC and, in general, SEM literature (see discussion of the topic in SEM context by, for example, Bentler, 2007; Hooper et al., 2008). In addition, it should be noticed that, with modeling the DE, the indicators are affected by the covariates twice. Directly, in the measurement model, and indirectly, through the latent variable. Resorting to a stepwise approach might indeed cause the effect in the later step be partially absorbed by the effect at the first step - see, for instance, Asparouhov & Muthén (2014). Nonetheless, in light of the results of our simulation study, not modeling the DE's can have worse effects on the parameters of interest.

Next to bias, we looked at coverage and variance of the estimators as well. All corrected models have approximately 95% coverage (given that the uncorrected models have biased estimates, looking at the coverage is less relevant). Comparing the corrected stepwise approaches to each other we can see that the augmented three-step approach has larger variability than the corrected two-step approach. These findings can also be seen in the real data application, where, using the augmented three-step approach, the same overall conclusion is reached as with the corrected one-step and two-step, but point estimates are further away from the other approaches.

A practical advice would be to use the two-steps approach if the direct effects are of substantive interest, whereas using the augmented three step is recommended when the direct effects are not of interest - but just a source of bias. An extra argument in favor of the three-step approach is the faster computation time (Bartolucci et al., 2015), which can make a difference for large datasets - especially with many time points, indicators and covariates.

In the first step of the augmented three-step approach, as we have pointed out earlier, the number of direct effects is likely to be misspecified. Considering that it is common, in applied research, to have a rough knowledge on where the direct effects can be located, the augmented three-step approach is likely to overspecify the direct effects, rather than underspecifying them. Whereby the latter, based on what we found to be the consequences of not modeling direct effects, can be harmful for parameter estimation, we found overspecification of the direct effects not to be an issue and to be safely dealt with by the absorbing latent variable of the augmented 3-step approach.

An issue related to the use of the augmented three-step approach is the choice of the number of classes for the absorbing latent variable. In principle, many states are more likely to approximate the supposedly continuous distribution of the noise generated in the overparametrized first step. This can results in issues of model identification and instability of the estimates. In our setups, we have tried several choices of P, and we found $P = 2 * S$ to be enough to approximate the noise distribution, still not being harmful to the model local identifiability. This heuristical solution worked for both the simulation study and the empirical application. On the other hand, we found that $P \leq S$ was not enough, therefore we suggest to pick values $P > S$. Although the P classes do not need to be substantively interpretable, whether P can be selected together with S, using information criteria (like BIC) based on the first step model, is indeed an interesting topic for future research.

In this paper we chose to treat the direct effects as known using the one-step and two-step approaches for the simplicity of exposition. However, often in applied research the exact location of the DEs is not know and it needs to be investigated. While we shortly discussed the issue of using residual statistics such as the L-BVR to identify direct effects, including this in the simulation study was outside the scope of this paper. As such we recommend for future research to investigate how direct effects can be best identified in the context of latent Markov modeling. In the context of simple LCA, the Expected Parameter Change (EPC) statistic and the BVR are successfully implemented (D. L. Oberski et al., 2013), the BVR is implemented also in the context of multilevel LCA (Nagelkerke et al., 2016). How they perform in the context of LM modeling is still to be investigated. Using the three-step approach, these statistics can not be used, so the safest and most lucrative option would be to have an overparametrized model, where also the resulting extra noise is modeled, as we propose in this paper.

Using both the two- and three-step corrected approaches, the direct effects need to be modeled in step one. It is still unknown the effect of adding covariates in the first step model on model selection, an issue that should be addressed by future research.

References

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

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

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

Bartolucci, F., & Farcomeni, A. (2009). A multivariate extension of the dynamic logit model for longitudinal data based on a latent markov heterogeneity structure. *Journal of the American Statistical Association*, *104*, 816-831.

Bartolucci, F., Farcomeni, A., & Pennoni, F. (2012). *Latent markov models for longitudinal data*. Chapman and Hall / CRC Press.

Bartolucci, F., Montanari, G., & Pandolfi, S. (2014). A comparison of some estimation methods for latent markov models with covariates. In M. Gilli, G. Gonzalez-Rodriguez, & A. Nieto-Reyes (Eds.), *Compstat 2014—proceedings in computational statistics* (pp. 423–431). Physica Verlag, Heidelberg, Germany.

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*, *41*(1), 164–171.

Bennink, M., Croon, M. A., & Vermunt, J. K. (2015). Stepwise latent class models for explaining group-level outcomes using discrete individual-level predictors. *Multivariate Behavioral Research*, *50*, 662-675.

Bentler, P. M. (2007). On tests and indices for evaluating structural models. *Personality and Individual Differences*, *42*, 825-829.

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

Di Mari, R., Oberski, D. L., & Vermunt, J. K. (2016). Bias-adjusted three-step latent markov modeling with covariates. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*, 649-660.

Hardin, J. W. (2002). The robust variance estimator for two-stage models. *The Stata Journal*, *2*, 253-266.

Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, *6*, 53-60.

Jackson, C. H., Sharples, L. D., Thompson, S. G., Duffy, S. W., & Couto, E. (2003). Multistate markov models for disease progression with classification error. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *52*, 193-209.

Kim, M., Vermunt, J. K., Bakk, Z., Jaki, T., & Van Horn, M. L. (2016). Modeling predictors of latent classes in regression mixture models. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*, 601-614.

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

McCutcheon, A. L. (1985). A latent class analysis of tolerance for nonconformity in the american public. *Public Opinion Quarterly*, *49*(4), 474–488.

Montanari, G. E., & Pandolfi, S. (2017). Evaluation of long-term health care services through a latent markov model with covariates. *Statistical Methods & Applications*. doi: 10.1007/s10260-017-0390-2

Murphy, K. M., & Topel, R. H. (1985). Estimation and inference in two-step econometric models. *Journal of Business and Economic Statistics*, *3*, 88-97.

Nagelkerke, E., Oberski, D. L., & Vermunt, J. K. (2016). Goodness-of-fit of multilevel latent class models for categorical data. *Sociological Methodology*, *46*, 252-282.

Nylund-Gibson, K., & Masyn, K. E. (2016). Covariates and mixture modeling: Results of a simulation study exploring the impact of misspecified effects on class enumeration. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*, 782-797.

Oberski, D. (2016). Beyond the number of classes: separating substantive from non-substantive dependence in latent class analysis. *Advances in Data Analysis and Classification*, *10*, 171-182.

Oberski, D. L., van Kollenburg, G. H., & Vermunt, J. K. (2013). A monte carlo evaluation of three methods to detect local dependence in binary data latent class models. *Advances in Data Analysis and Classification*, *7*, 267-279.

Poulsen, C. S. (1990). Mixed markov and latent markov modelling applied to brand choice behaviour. *International Journal of Research in Marketing*, *7*, 5-19.

Vermunt, J., & Magidson, J. (2004). Factor analysis with categorical indicators: A comparison between traditional and latent class approaches. In L. van der Ark, M. Croon, & K. Sijtsma (Eds.), *New developments in categorical data analysis for the social and behavioral sciences* (p. 41-63). Mahwah, NJ: Erlbaum.

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

Vermunt, J. K., Langeheine, R., & B ockenholt, 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.

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

Vermunt, J. K., & Magidson, J. (2016). Technical guide for latent gold 5.1: Basic, advanced, and syntax. *Belmont, MA: Statistical Innovations Inc.*.

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

Wiggins, L. M. (1973). *Panel analysis: latent probability models for attitude and behaviour processes*. Elsevier, Amsterdam.

Zhu, Y., Steele, F., & Moustaki, I. (2017). A general 3-step maximum likelihood approach to estimate the effects of multiple latent categorical variables on a distal outcome. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(5), 643-656.

Zucchini, W., MacDonald, I. L., & Langrock, R. (2016). *Hidden markov models for time series: an introduction using r*. Chapman and Hall / CRC Press.

Appendix A

Latent GOLD syntax for the real data example for the two-step approach

Step one model including the direct effects

```
output parameters=first standarderrors bivariateresiduals;
variables
independent cohort, education;
dependent spkath, spkcom, spkhomo, spkmil, spkrac, SPKMSLM;
latent
Cluster nominal 3;
equations
Cluster ← 1;
spkath ← 1 + Cluster;
spkcom ← 1 + Cluster;
spkhomo ← 1 + Cluster+ cohort;
spkmil ← 1 + Cluster;
```

spkrac ← 1 + Cluster+ education;
SPKMSLM ← 1 + Cluster; { # In case the DE's ares not modeled leave out the effect of cohort and education on the indicators of interest };

Step two model

output parameters=first standarderrors ;
variables
caseid id;
dependent spkath, spkcom, spkhomo, spkmil, spkrac, SPKMSLM;
independent education, cohort;
latent State dynamic nominal 3;
equations
State[=0] ← 1+education+ cohort ;
State ← 1+ State[-1] +education+ cohort ;
spkath ←(a)1 + (a1) State;
spkcom ← (b)1 + (b1)State;
spkhomo ← (c)1 + (c1)State+ (c2)cohort;
spkmil ← (d)1 + (d1)State;
spkrac ← (e)1 + (e1)State+ (e2)education;
SPKMSLM ← (f)1 + (f1)State;
a a1 b b1 c c1 c2 d d1 e e1 e2 f f1=
{ # place here the fixed value parameters estimated in step one. };
{ # In case the DE's ares not modeled leave out the effect of cohort and education on the indicators of interest };

Appendix B

Latent GOLD syntax for the real data example for the three step approach

Step one & two model

output parameters=first standarderrors ;
outfile 'datastep3.sav' classification keep id, TIME;
variables dependent spkath, spkcom, spkhomo, spkmil, spkrac, SPKMSLM;
independent education, cohort;
latent Cluster nominal 3, class nominal 6;
equations
Cluster ← 1;
class ← 1;
spkath ← 1 + Cluster;
spkcom ←1 + Cluster;
spkhomo ← 1 + Cluster+ class+ education;
spkmil ← 1 + Cluster;
spkrac ← 1 + Cluster+class+cohort;
SPKMSLM ← 1 + Cluster + class+cohort; { # In case the DE's ares not modeled leave out the second latent variable and the effect of cohort and education on the indicators of interest };

Step three model

step3 modal ml;
output parameters=first standarderrors=robust;
variables
caseid id;

independent education, cohort;
latent Cluster dynamic nominal posterior = (Cluster#1 Cluster#2 Cluster#3);
equations
Cluster[=0] ← 1+education+ cohort ;
Cluster ← 1+ Cluster[-1] +education+ cohort ;

Appendix C

Detailed look on the SE and SD obtained with the different estimators in all the conditions for one selected parameter

Table A1

Simulation Results - Average estimated SE, SD for all six simulation conditions for β_{2Z_1}. One Direct Effect.

Sample Size	100		500		1000	
	SD	SE	SD	SE	SD	SE
			Moderate separation			
FIMLcorr	0.591	0.571	0.243	0.248	0.165	0.175
FIMLuncorr	0.591	0.570	0.244	0.247	0.164	0.175
2STEPuncor	0.585	0.570	0.242	0.248	0.164	0.175
2STEPcorr	0.596	0.577	0.228	0.249	0.179	0.175
3STEP	0.611	0.603	0.256	0.258	0.167	0.182
3STEPaug	0.607	0.617	0.261	0.260	0.185	0.183
			Large separation			
FIMLcorr	0.515	0.533	0.240	0.234	0.158	0.165
FIMLuncorr	0.515	0.533	0.240	0.234	0.158	0.165
2STEPuncor	0.511	0.534	0.240	0.234	0.158	0.165
2STEPcorr	0.503	0.533	0.234	0.234	0.161	0.165
3STEP	0.523	0.548	0.244	0.238	0.162	0.168
3STEPaug	0.534	0.553	0.247	0.239	0.170	0.168

Table A2

Simulation Results - Average estimated SE, SD for all six simulation conditions for β_{2Z_1}. Two Direct Effects.

Sample Size	100		500		1000	
	SD	SE	SD	SE	SD	SE
			Moderate separation			
FIMLcorr	0.622	0.577	0.244	0.249	0.173	0.175
FIMLuncorr	0.616	0.575	0.245	0.249	0.172	0.175
2STEPuncor	0.604	0.577	0.245	0.249	0.173	0.175
2STEPcorr	0.608	0.579	0.244	0.249	0.173	0.175
3STEP	0.626	0.609	0.255	0.259	0.185	0.182
3STEPaug	0.620	0.615	0.258	0.264	0.193	0.186
			Large separation			
FIMLcorr	0.551	0.535	0.227	0.234	0.159	0.165
FIMLuncorr	0.55	0.535	0.227	0.234	0.159	0.165
2STEPuncor	0.543	0.537	0.227	0.234	0.158	0.165
2STEPcorr	0.544	0.537	0.226	0.234	0.159	0.165
3STEP	0.548	0.549	0.231	0.238	0.161	0.168
3STEPaug	0.555	0.554	0.236	0.241	0.162	0.170

Table A3

Simulation Results - Average estimated SE, SD for all six simulation conditions for β_{2Z_1}. Three Direct Effects.

Sample Size	100		500		1000	
	SD	SE	SD	SE	SD	SE
			Moderate separation			
FIMLcorr	0.559	0.575	0.244	0.249	0.168	0.176
FIMLuncorr	0.563	0.575	0.244	0.249	0.171	0.175
2STEPuncor	0.556	0.577	0.242	0.249	0.170	0.175
2STEPcorr	0.554	0.579	0.243	0.250	0.168	0.176
3STEP	0.587	0.609	0.248	0.260	0.18	0.183
3STEPaug	0.569	0.611	0.254	0.265	0.187	0.187
			Large separation			
FIMLcorr	0.546	0.535	0.247	0.234	0.163	0.165
FIMLuncorr	0.545	0.535	0.248	0.234	0.162	0.165
2STEPuncor	0.539	0.537	0.247	0.234	0.162	0.165
2STEPcorr	0.539	0.537	0.247	0.234	0.162	0.165
3STEP	0.544	0.550	0.249	0.239	0.165	0.168
3STEPaug	0.550	0.555	0.251	0.242	0.165	0.170