TWO-STEP ESTIMATION OF MODELS BETWEEN LATENT CLASSES AND EXTERNAL VARIABLES

ZSUZSA BAKK

LEIDEN UNIVERSITY

JOUNI KUHA

LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

We consider models which combine latent class measurement models for categorical latent variables with structural regression models for the relationships between the latent classes and observed explanatory and response variables. We propose a two-step method of estimating such models. In its first step, the measurement model is estimated alone, and in the second step the parameters of this measurement model are held fixed when the structural model is estimated. Simulation studies and applied examples suggest that the two-step method is an attractive alternative to existing one-step and three-step methods. We derive estimated standard errors for the two-step estimates of the structural model which account for the uncertainty from both steps of the estimation, and show how the method can be implemented in existing software for latent variable modelling.

Key words: latent variables, mixture models, structural equation models, pseudo-maximum likelihood estimation.

1. Introduction

Latent class analysis is used to classify objects into categories on the basis of multiple observed characteristics. The method is based on a model where the observed variables are treated as measures of a latent variable which has some number of discrete categories or *latent classes* (Lazarsfeld & Henry, 1968; Goodman, 1974; Haberman, 1979; see McCutcheon, 1987 for an overview). This has a wide range of applications in psychology, other social sciences, and elsewhere. For example, latent class analysis was used to identify types of substance abuse among young people by Kam (2011), types of music consumers by Chan and Goldthorpe (2007), and patterns of workplace bullying by Einarsen, Hoel, and Notelaers (2009).

In many applications, the interest is not just in clustering into the latent classes but in using these classes in further analysis with more complex models. Such extensions include using observed covariates (explanatory variables) to predict latent class membership and using the latent class as a covariate for other outcomes. For instance, in our illustrative examples we examine how education and birth cohort predict tolerance for nonconformity as classified by latent class analysis, and how latent classes of perceived psychological contract between employer and employee predict the employee's feelings of job insecurity.

Models like these have two main components: the *measurement model* for how the latent classes are measured by their observed indicators, and the *structural model* for the relationships between the latent classes and other explanatory or response variables. Different approaches may

be used to fit these models, differing in how the structural and measurement models are estimated and whether they are estimated together or in separate steps. In this article, we propose a new "two-step" method of estimating such models and show that it is an attractive alternative to existing "one-step" and "three-step" methods.

In the *one-step method* of estimation, both parts of the model are estimated at the same time, to obtain maximum likelihood (ML) estimates for all of their parameters (see e.g. Clogg, 1981, Dayton & Macready, 1988, Hagenaars, 1993, Bandeen-Roche, Miglioretti, Zeger, & Rathouz, 1997, and Lanza, Tan, & Bray, 2013; this is also known as full-information ML, or FIML, estimation). Although this approach is efficient and apparently natural, it also has serious defects (see e.g. the discussions in Croon, 2002, Vermunt, 2010, and Asparouhov & Muthén, 2014). These arise because the whole model is always refitted even when only one part of it is changed. Practically, this can make the estimation computationally demanding, especially if we want to fit many models to compare structural models with multiple variables. The more disturbing problem with the one-step approach, however, is not practical but conceptual: every change in the structural model—for example adding or removing covariates—affects also the measurement model and thus in effect changes the definition of the latent classes, which in turn distorts the interpretation of the results of the analysis. This problem is not merely hypothetical but can in practice occur to an extent which can render comparisons of estimated structural models effectively meaningless. One of our applied examples in this article, which is discussed in Sect. 4.1, provides an illustration of this phenomenon.

"Stepwise" methods avoid the problems of the one-step approach by separating the estimation of the different parts of the model into distinct steps of the analysis. Most existing applications of this idea to latent class analysis are different versions of the *three-step method*. This involves (1) estimating the measurement model alone, using only data on the indicators of the latent classes, (2) assigning predicted values of the latent classes to the units of analysis based on the model from step 1, and (3) estimating the structural model with the assigned values from step 2 in the role of the latent classes. The most common version of this is the *naive three-step method* where the values assigned in step 2 are treated as known variables in step 3. In this as in all the stepwise methods, the first-step modelling may even be done by different researchers or with different data than the subsequent steps.

The naive three-step method has the flaw that the values assigned in its second step are not equal to the true values of the latent classes as defined by the first step. This creates a measurement error (misclassification) problem which means that the third step will yield biased estimates of the structural model (Croon, 2002). The misclassification can be allowed for and the biases corrected by using *bias-adjusted three-step methods* (Bolck, Croon, & Hagenaars, 2004; Vermunt, 2010; Bakk, Tekle, & Vermunt, 2013; Asparouhov & Muthén, 2014) which have been developed in recent years and which are now also implemented in two mainstream software packages for latent class analysis, Latent GOLD (Vermunt & Magidson, 2005, 2016) and Mplus (Muthén & Muthén, 2017). However, applied researchers who are unfamiliar with the correction methods, or who are using other software packages, will still most often be using the naive three-step approach.

In this paper, we propose an alternative *two-step method* of estimation. Its first step is the same as in the three-step methods, that is fitting the latent class measurement model on its own. In the second and final step, we then maximize the joint likelihood (i.e. the likelihood which is also used in the one-step method), but with the parameters of the measurement model and of exogenous latent variables (if any) fixed at their estimated values from the first step, so that only the parameters of the rest of the structural model are estimated in the second step. This proposal is rooted in the realization that the essential feature of a stepwise approach is that the measurement model is estimated separately, not that there needs to be an explicit classification step. This is especially important because the classification error of the three-step method is introduced in its second step. So by eliminating this step, we eliminate the circular problem of introducing an error

that we then need to correct for later on. As a result, the two-step method is more straightforward and easier to understand than the bias-adjusted three-step methods.

This approach was suggested as a possibility already by Bandeen-Roche et al. (1997, p. 1384). Xue and Bandeen-Roche (2002) developed it in full, in their case for structural models with the latent class as the response variable, and motivated by applications where the first step was based on a much larger sample than the second. It was also used by Bartolucci, Montanari, and Pandolfi (2014) for latent Markov models for longitudinal data. We build on and extent these previous proposals and describe two-step modelling and its properties as a general method for latent class analysis. As already noted by Xue and Bandeen-Roche (2002), it can be motivated as an instance of two-stage pseudo-ML estimation (Gong & Samaniego, 1981). The general theory of such estimation shows that the two-step estimates of the parameters of the structural model are consistent, and it provides asymptotic variance estimates which correctly allow also for the uncertainty in the estimates from the first step. Software which can carry out one-step estimation can also be used to implement the two-step method. Our simulations suggest that the two-step estimates are typically only slightly less efficient than the one-step estimates, and a little more efficient than the bias-adjusted three-step estimates.

Although we focus in this article on latent class models, the conceptual issues and the methods that we describe apply also to other latent variable models (we discuss this briefly further in Sect. 5). In particular, they are also relevant for structural equation models (SEMs) where both the latent variables and their indicators are treated as continuous variables (see e.g. Bollen, 1989). There the most commonly used methods are one-step (standard SEMs) and naive three-step estimation (using factor scores as derived variables). For some models, it is possible to assign factor scores in such a way that the bias of the naive three-step approach is avoided (Skrondal & Laake, 2001; comparable methods have been proposed for item response theory models by Lu & Thomas, 2008, and for latent class models by Petersen, Bandeen-Roche, Budtz-Jørgensen, & Groes Larsen, 2012), and bias-corrected three-step methods can also be developed (Croon, 2002; Devlieger, Mayer, & Rosseel, 2016), but these approaches are less often used in practice. Another stepwise approach for SEMs is two-stage least squares (2SLS) estimation, different versions of which have been proposed by Jöreskog and Sörbom (1986), Lance, Cornwell, and Mulaik (1988) and Bollen (1996). It uses the ideas of instrumental variable estimation and is quite different in form to our two-step method. In the closely related context of generalized linear models with continuous covariates measured with error, Skrondal and Kuha (2012) proposed a two-step pseudo-ML method which is essentially analogous to the one which is described here for latent class models (it uses a slightly different split of parameters between steps one and two).

The conceptual disadvantages of the one-step method were discussed in the context of SEMs already by Burt (1976, 1973). He introduced the idea of "interpretational confounding" which arises when the variables that a researcher uses to interpret a latent variable differ from the variables which actually contribute to the estimation of its measurement model. As a way of avoiding such confounding, Burt proposed a stepwise approach which was two-step estimation in the same sense that we describe here. Subsequent literature has, however, made little use of this proposal, even when it has drawn on Burt's ideas otherwise. In particular, stepwise thinking is now much more commonly applied to model selection rather than estimation—in other words, the *form* of the measurement model is selected separately, but the parameters of this measurement model and any structural models are then estimated together using one-step estimation (Anderson & Gerbing, 1988). It is likely that in the large SEM literature there are individual instances of the use of two-step estimation (one example is Ping, 1996), but they are clearly not widespread. There appear to be no systematic theoretical expositions of two-step estimation of the kind that is offered in this article.

The model setting and the method of two-step estimation are introduced in Sect. 2 below, followed in Sect. 3 by a simulation study where we compare it to the existing one-step and

three-step approaches. We then illustrate the method in two applied examples in Sect. 4 and give concluding remarks in Sect. 5.

2. Two-Step Estimation of Latent Class Models with External Variables

2.1. The Variables and the Models

Let X be a latent variable, $\mathbf{Y} = (Y_1, \ldots, Y_K)$ observed variables which are treated as measures (indicators) of X, and $\mathbf{Z} = (\mathbf{Z}_p, Z_o)$ observed variables where $\mathbf{Z}_p$ are covariates (predictors, explanatory variables) for X and Z_o is a response variable to X and $\mathbf{Z}_p$. We take X and Z_o to be univariate for simplicity of presentation, but this can easily be relaxed as discussed later. Here X is a categorical variable with C categories (latent classes) $c = 1, \ldots, C$, and each of the indicators Y_k is also categorical, with R_k categories for $k = 1, \ldots, K$. Suppose we have a sample of data for n units such as survey respondents, so that the observed data consist of $(\mathbf{Z}_i, \mathbf{Y}_i)$ for $i = 1, \ldots, n$, while X_i remain unobserved.

We denote marginal density functions and probabilities by $p(\cdot)$ and conditional ones by $p(\cdot|\cdot)$. The *measurement model* for X_i for a unit i is given by

$$p(\mathbf{Y}_i|X_i, \mathbf{Z}_i) = p(\mathbf{Y}_i|X_i = c) = \prod_{k=1}^{K} p(Y_{ik}|X_i = c) = \prod_{k=1}^{K} \prod_{r=1}^{R_k} \pi_{kcr}^{I(Y_{ik}=r)} \tag{1}$$

for $c = 1, \ldots, C$, where π_{kcr} are probability parameters and $I(Y_{ik} = r) = 1$ if unit i has response r on measure k, and 0 otherwise. This is the measurement model of the *latent class model* with K categorical indicator variables for C latent classes. It is assumed here that $\mathbf{Y}_i$ are conditionally independent of $\mathbf{Z}_i = (\mathbf{Z}_{pi}, Z_{oi})$ given X_i (i.e. that $\mathbf{Y}_i$ are purely measures of the latent X_i and there are no direct effects from other observed variables $\mathbf{Z}_i$ to $\mathbf{Y}_i$) and that the indicators Y_{ik} are conditionally independent of each other given X_i. These are standard assumptions of basic latent class analysis.

The *structural* model $p(\mathbf{Z}_i, X_i) = p(\mathbf{Z}_{pi}) p(X_i|\mathbf{Z}_{pi}) p(Z_{oi}|\mathbf{Z}_{pi}, X_i)$ specifies the joint distribution of $\mathbf{Z}_i$ and X_i. Then, $p(\mathbf{Z}_i, X_i, \mathbf{Y}_i) = p(\mathbf{Z}_i, X_i) p(\mathbf{Y}_i|X_i)$, and the distribution of the observed variables is obtained by summing this over the latent classes of X_i to get

$$p(\mathbf{Z}_i, \mathbf{Y}_i) = p(\mathbf{Z}_{pi}) \sum_{c=1}^{C} \left[p(X_i = c|\mathbf{Z}_{pi}) \, p(Z_{oi}|\mathbf{Z}_{pi}, X_i = c) \prod_{k=1}^{K} p(Y_{ik}|X_i = c) \right]. \tag{2}$$

This model thus combines a latent class measurement model for X_i with a structural model for the associations between X_i and observed covariates $\mathbf{Z}_{pi}$ and/or response variables Z_{oi}.

Substantive research questions typically focus on those parts of the structural model which involve X, so the primary goal of the analysis is to estimate $p(X_i|\mathbf{Z}_{pi})$ and/or $p(Z_{oi}|\mathbf{Z}_{pi}, X_i)$. The measurement model is then of lesser interest, but it too needs to be specified and estimated correctly to obtain valid estimates for the structural model, not least because the measurement model provides the definition and interpretation of X_i. The marginal distribution $p(\mathbf{Z}_{pi})$ can be dropped and the estimation done conditionally on the observed values of $\mathbf{Z}_{pi}$.

For simplicity of illustrating the methods in specific situations, we will focus on structural models where either $\mathbf{Z}_p$ or Z_o is absent. These cases will be considered in our simulations in Sect. 3 and the examples in Sect. 4. We thus consider first the case where there is no Z_o and the object of interest is $p(X_i = c|\mathbf{Z}_{pi})$, the model for how the probabilities of the latent classes depend on observed covariates $\mathbf{Z}_p$. This is specified as the multinomial logistic model

$$p(X_i = c|\mathbf{Z}_{pi}) = \frac{\exp(\beta_{0c} + \boldsymbol{\beta}_c\mathbf{Z}'_{pi})}{\sum_{s=1}^{C}\exp(\beta_{0s} + \boldsymbol{\beta}_s\mathbf{Z}'_{pi})} \tag{3}$$

for $c = 1, \ldots, C$, and $(\beta_{01}, \boldsymbol{\beta}_1) = \mathbf{0}$ for identifiability (and $\boldsymbol{\beta}_c$ and $\mathbf{Z}_{pi}$ are taken to be row vectors). Second, we consider the case where there is no $\mathbf{Z}_p$ and the object of interest is $p(Z_{oi}|X_i)$, a regression model for an observed response variable Z_o given latent class membership. Here $p(X_i = c|\mathbf{Z}_{pi}) = p(X_i = c)$, the explanatory variables in $p(Z_{oi}|X_i)$ are dummy variables for the latent classes $c = 2, \ldots, C$, and the form of this model depends on the type of Z_o. In our simulations and applied example, Z_o is a continuous variable and the model for it is a linear regression model.

2.2. Existing Approaches: The One-Step and Three-Step Methods

Let $\boldsymbol{\theta} = (\boldsymbol{\pi}, \boldsymbol{\psi}_p, \boldsymbol{\psi}_o)$ denote the parameters of the joint model, where $\boldsymbol{\pi}$ are the parameters of the measurement model, $\boldsymbol{\psi}_p$ of the structural model for X given $\mathbf{Z}_p$ (or just the probabilities $p(X_i = c)$, if there are no $\mathbf{Z}_p$), and $\boldsymbol{\psi}_o$ of the structural model for Z_o (if any). If the units i are independent, the log likelihood for $\boldsymbol{\theta}$ is $\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n}\log p(Z_{oi}, \mathbf{Y}_i|\mathbf{Z}_{pi})$, obtained from (2) by omitting the contribution from $p(\mathbf{Z}_{pi})$. Maximizing $\ell(\boldsymbol{\theta})$ gives maximum likelihood (ML) estimates of all of $\boldsymbol{\theta}$. These are the *one-step estimates* of the parameters. They are most conveniently obtained using established software for latent variable modelling, currently in particular Latent GOLD or Mplus. These software typically use the EM algorithm, a quasi-Newton method, or a combination of them, to maximize the log likelihood. They also provide other estimation facilities which are important for complex latent variable models, such as automatic implementation of multiple starting values.

Stepwise methods of estimation begin instead with the more limited log likelihood $\ell_1(\boldsymbol{\rho}, \boldsymbol{\pi}) = \sum_{i=1}^{n}\log p(\mathbf{Y}_i)$, where

$$p(\mathbf{Y}_i) = \sum_{c=1}^{C}\left[p(X_i = c)\prod_{k=1}^{K}p(Y_{ik}|X_i = c)\right], \tag{4}$$

$\boldsymbol{\pi}$ are the same measurement parameters (response probabilities) as defined above, and $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_C)$ with $\rho_c = p(X_i = c) = \int p(\mathbf{Z}_{pi})p(X_i = c|\mathbf{Z}_{pi})\,d\mathbf{Z}_{pi}$; thus, $\boldsymbol{\rho}$ are the same as $\boldsymbol{\psi}_p$ if there are no covariates $\mathbf{Z}_p$ but not otherwise. Expression (4) defines a standard latent class model without covariates or response variables $\mathbf{Z}$. In step 1 of all of the stepwise methods, we maximize $\ell_1(\boldsymbol{\rho}, \boldsymbol{\pi})$ to obtain ML estimates of the parameters of this model. This step 1 log likelihood can also be based on a partially or completely different set of observations than $\ell(\boldsymbol{\theta})$; this possibility is discussed further in Sect. 2.3.

Since step 1 model gives estimates of $p(X_i = c)$ and $p(\mathbf{Y}_i|X_i = c)$, it also implies estimates of the probabilities $p(X_i = c|\mathbf{Y}_i)$ of latent class membership given observed response patterns $\mathbf{Y}_i$. In step 2 of a three-step method, these conditional probabilities are used in some way to assign to each unit i a value $\tilde{c}_i$ of a new variable $\tilde{X}_i$ which will be used as a substitute for X_i. The most common choice is the "modal" assignment, where $\tilde{c}_i$ is the single value for which $p(X_i = c|\mathbf{Y}_i)$ is highest. In *naive three-step estimation*, step 3 then consists of using $\tilde{X}_i$ as an observed variable in the place of X_i when estimating the structural models for the associations between X_i and $\mathbf{Z}_i$, to obtain naive three-step estimates of the parameters of interest $\boldsymbol{\psi}_p$ and/or $\boldsymbol{\psi}_o$. These estimates are, however, generally biased, because of the misclassification error induced by the fact that $\tilde{X}_i$ are not equal to X_i. It is important to note that this bias arises not just from modal assignment but from any step 2 assignment whose misclassification is not subsequently allowed for; this includes

even methods where each unit is assigned to every latent class with fractional weights which are proportional to $p(X_i = c|\mathbf{Y}_i)$ (Dias & Vermunt, 2008; Bakk et al., 2013).

Bias-adjusted three-step methods remove this problem of the naive methods. Their basic idea is to use the estimated misclassification probabilities $p(\tilde{X}_i = \tilde{c}_i | X_i = c_i)$ of the values assigned in step 2 to correct for the misclassification bias. The two main approaches for doing this are the "BCH" method proposed by Bolck et al. (2004) and extended by Vermunt (2010) and Bakk et al. (2013), and the "ML" method proposed by Vermunt (2010) and extended by Bakk et al. (2013) (see also Asparouhov & Muthén, 2014). Step 3 of the BCH method uses $\tilde{X}_i$ explicitly in place of X_i, in the same way as in naive three-step estimation, but with weighting used to adjust for the misclassification. In contrast, step 3 of the ML method involves maximizing a log likelihood which has the same form as $\ell(\boldsymbol{\theta})$, except that $p(\mathbf{Y}_i|X_i)$ is replaced with $p(\tilde{X}_i|X_i)$ and this is fixed at its estimate from step 2 (it is thus closer in spirit to the two-step method, whose second and final step will involve similar fixing, but applied directly to $p(\mathbf{Y}_i|X_i)$). Both of these adjusted three-step methods are available in Latent GOLD and Mplus, while in other software additional programming would be required.

2.3. The Proposed Two-Step Method

We propose a *two-step method* of estimation. Its first step is the same as in the three-step methods, that is estimating the latent class model (4) without covariates or response variables $\mathbf{Z}$. Some or all of the parameter estimates from this model are then passed on to the second step and treated as fixed there, while the rest of the parameters of the full model are estimated.

Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ denote the decomposition of $\boldsymbol{\theta}$ into those parameters that will be estimated in step 1 ($\boldsymbol{\theta}_1$) and those that will be estimated in step 2 ($\boldsymbol{\theta}_2$). There are two possibilities regarding what we will include in $\boldsymbol{\theta}_1$ (these two situations are also represented graphically in Fig. 1). If there are any covariates $\mathbf{Z}_p$, then $\boldsymbol{\theta}_1 = \boldsymbol{\pi}$, i.e. it includes only the parameters of the measurement model (and estimates of $\boldsymbol{\rho}$ from step 1 will be discarded before step 2). If there are no $\mathbf{Z}_p$, then $\boldsymbol{\theta}_1 = (\boldsymbol{\pi}, \boldsymbol{\psi}_p)$, i.e. it includes also the probabilities $\boldsymbol{\psi}_p = \boldsymbol{\rho}$ of the marginal distribution of X. The logic of this second choice is that if X is not a response variable to any $\mathbf{Z}_p$, we can treat it as an exogenous variable whose distribution can also be estimated from step 1 and then treated as fixed when we proceed in step 2 to the estimation of models conditional on X. Thus, $\boldsymbol{\theta}_2$ includes either all the parameters $(\boldsymbol{\psi}_p, \boldsymbol{\psi}_o)$ of the structural model, or all of them except those of an exogenous X.

Denoting the estimates of $\boldsymbol{\theta}_1$ from step 1 by $\tilde{\boldsymbol{\theta}}_1$, in step 2 we use the log likelihood $\ell_2(\tilde{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2)$, which is $\sum_{i=1}^{n} \log p(Z_{oi}, \mathbf{Y}_i | \mathbf{Z}_{pi})$ evaluated at $\boldsymbol{\theta}_1 = \tilde{\boldsymbol{\theta}}_1$ and treated as a function of $\boldsymbol{\theta}_2$ only. Maximizing this with respect to $\boldsymbol{\theta}_2$ gives the two-step estimate of these parameters, which we denote by $\tilde{\boldsymbol{\theta}}_2$.

This procedure achieves the aims of stepwise estimation, because the measurement model is held fixed when (all or most of) the structural model is estimated. If we change the structural model, $\tilde{\boldsymbol{\theta}}_1$ remains the same and only step 2 is done again (or even if we do run both steps again, $\tilde{\boldsymbol{\theta}}_1$ will not change). This would be the case, for example, if we wanted to compare models with different explanatory variables $\mathbf{Z}_p$ for the same latent class variable X.

Another useful aspect of separating the estimation of the measurement and structural models is that the estimates $\tilde{\boldsymbol{\theta}}_1$ and $\tilde{\boldsymbol{\theta}}_2$ may be obtained using different samples. A common example of this is that some observations which are used for step 1 may be omitted in step 2 because of missing data in $\mathbf{Z}$. A more dramatic instance occurs when, because of resource constraints, $\mathbf{Z}$ is measured for a subset of units only, so that step 1 is based on a much larger sample (for example, this was a key motivation of two-step estimation in the application considered by Xue & Bandeen-Roche, 2002). Conversely, we might sometimes decide to keep $\tilde{\boldsymbol{\theta}}_1$ unchanged even when new data on $\mathbf{Z}$ become available, so that step 2 may be based on a larger sample (or even a completely different
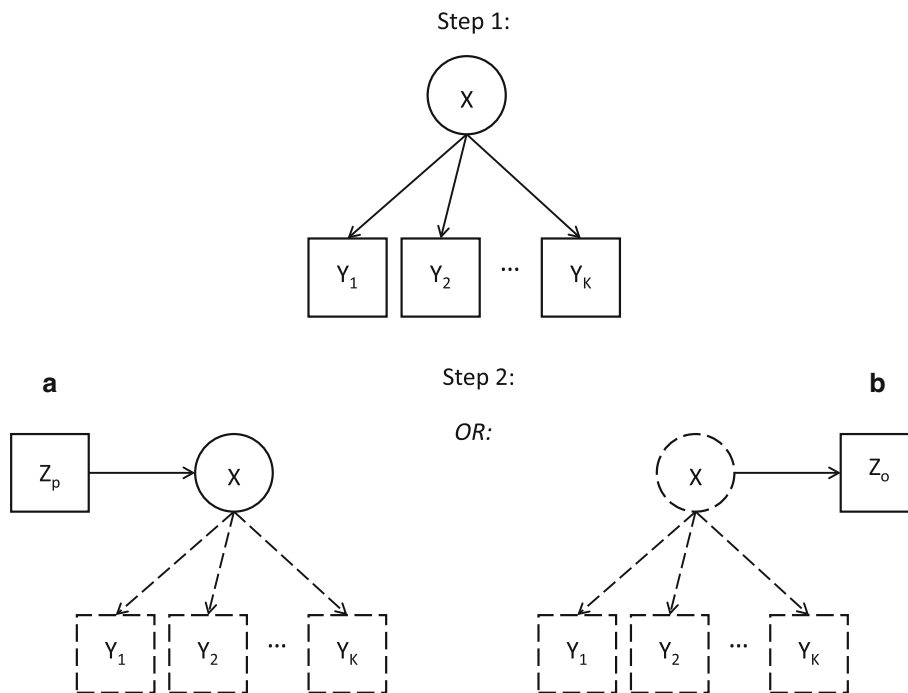
Step 1:

Step 2:

OR:

a

b

FIGURE 1.

Graphical representation of the two-step method of latent class analysis with latent class variable X measured by indicators $Y_1, Y_2, \ldots, Y_K$. Two specific structural models are represented, **a** with only covariates Z_p for X and **b** with only response variables Z_o for it. In step 2, the dashed lines represent those parts of the model which are held fixed at their estimates from step 1.

sample) than step 1. In all of these cases, the two-step estimate $\tilde{\boldsymbol{\theta}}_2$ will be consistent for $\boldsymbol{\theta}_2$ as long as the data are such that one-step estimates obtained from step 2 sample would also be consistent (for example, that any missing data there are ignorable for likelihood inference) and that even if step 1 and step 2 samples are different they both represent populations where the true value of $\boldsymbol{\theta}_1$ is the same.

Although we focus here on the case of a single X for simplicity, the idea of the two-step method extends naturally also to more complex situations. For instance, suppose that there are two latent class variables X_1 and X_2 with separate sets of indicators $\mathbf{Y}_1$ and $\mathbf{Y}_2$, and the structural model is of the form $p(X_1)p(Z_1|X_1)p(X_2|Z_1, X_1)p(Z_2|X_1, Z_1, X_2)$. In step 1, we would then estimate two separate latent class models, one for X_1 and one for X_2 (and both again without $\mathbf{Z} = (Z_1, Z_2)$). Step 1 parameters $\boldsymbol{\theta}_1$ would be the measurement probabilities of X_1 and X_2 and the parameters of $p(X_1)$, and step 2 parameters would be those of the rest of the structural model apart from $p(X_1)$.

2.4. Properties and Implementation of Two-Step Estimators

Two-step estimation in latent class analysis is an instance of a general approach to estimation where the parameters of a model are divided into two sets and estimated in two stages. The first set is estimated in the first step by some consistent estimators, and the second set of parameters is then estimated in the second step with the estimates from the first step treated as known. When the second step is done by maximizing a log likelihood, as is the case here, this is known as

pseudo-maximum likelihood (PML) estimation (Gong & Samaniego, 1981). The properties of our two-step estimators can be derived from the general PML theory.

Such two-stage estimators are consistent and asymptotically normally distributed under very general regularity conditions (see Gourieroux & Monfort, 1995, Sections 24.2.4 and 24.2.2). In our situation, these conditions are satisfied because the one-step estimator $\hat{\theta}$ and step 1 estimator $\tilde{\theta}_1$ of the two-step method are both ML estimators, of the joint model and the simple latent class model (4), respectively, and because the models are such that θ_1 and θ_2 can vary independently of each other.

Suppose that step 2 is based on n observations and step 1 on n_1 observations (which may be different, as discussed above). Let the Fisher information matrix for θ in the joint (one-step) model be

$$\mathcal{I}(\theta^*) = \begin{bmatrix} \mathcal{I}_{11} \\ \mathcal{I}'_{12} \ \mathcal{I}_{22} \end{bmatrix}$$

where θ^* denotes the true value of θ and the partitioning corresponds to θ_1 and θ_2. The asymptotic variance matrix of the one-step estimator $\hat{\theta}$ is thus $\mathbf{V}_{ML} = \mathcal{I}^{-1}(\theta^*)/n$, which is estimated by $\hat{\mathbf{V}}_{ML} = \mathcal{I}^{-1}(\hat{\theta})/n$. Similarly, let Σ_{11}/n_1 be the asymptotic variance matrix of step 1 estimator $\tilde{\theta}_1$ of the two-step method, obtained from the Fisher information matrix for model (4) and evaluated at the true values (ρ^*, π^*) of its parameters; this is estimated by substituting step 1 estimates for these parameters. The asymptotic variance matrix of the two-step estimator $\tilde{\theta}_2$ is then $\mathbf{V}/n$, where

$$\mathbf{V} = \mathcal{I}_{22}^{-1} + \mathcal{I}_{22}^{-1} \mathcal{I}'_{12} \left[(n/n_1)\Sigma_{11} \right] \mathcal{I}_{12} \mathcal{I}_{22}^{-1} \equiv \mathbf{V}_2 + \mathbf{V}_1 \tag{5}$$

where the (n/n_1) adjusts for the possibly different sample sizes in the two steps (see Xue & Bandeen-Roche, 2002). Here $\mathbf{V}_2$ describes the variability in $\tilde{\theta}_2$ if step 1 parameters θ_1 were actually known, and $\mathbf{V}_1$ the additional variability arising from the fact that θ_1 are not known but estimated by $\tilde{\theta}_1$. Comparable methods for bias-adjusted three-step estimators, which also allow for both of these sources of variation, have been proposed by Bakk, Oberski, and Vermunt (2014a). $\mathbf{V}$ is estimated by $\hat{\mathbf{V}} = \hat{\mathbf{V}}_2 + \hat{\mathbf{V}}_1$, obtained by substituting $\tilde{\theta} = (\tilde{\theta}_1, \tilde{\theta}_2)$ for θ^* in $\mathcal{I}_{22}$ and $\mathcal{I}_{12}$ in (5), evaluated using the n observations used for step 2, and the estimate from step 1 for Σ_{11}. The estimated variance matrix $\hat{\mathbf{V}}/n$ is then used to calculate confidence intervals for the parameters in θ_2 and Wald test statistics for them.

The standard errors that are routinely displayed by the software when we fit step 2 model are based on $\hat{\mathbf{V}}_2$ only. Because they omit the contribution from $\hat{\mathbf{V}}_1$, these standard errors will underestimate the full uncertainty in $\tilde{\theta}_2$. In the simulations of Sect. 3, we examine the magnitude of this underestimation in different circumstances. The results suggest that the contribution from step 1 uncertainty can be substantial and that it can be safely ignored only if the measurement model is such that $\mathbf{Y}$ are very strong measures of X.

As noted in the previous section, if the joint model had more than one latent class variable X, in the first step we would propose to estimate the latent class models for each of these variables separately. Even then, the estimated parameters of these models would be correlated, because they are estimated using data for the same units. An estimate which takes this into account can be obtained from the theory of estimating equations, using only the score functions and information matrices for the separate models (see e.g. Cameron & Trivedi, 2005, Section 5.4). A still simpler approach would be to approximate Σ_{11} by a block-diagonal matrix, with the blocks being the variance matrices for the distinct latent class models. This would ignore the correlations between these blocks of step 1 parameter estimates and would thus imply some misspecification of the resulting form of $\mathbf{V}_1$, but we might expect the effect of this misspecification to be relatively small.

In Appendix, we outline how the quantities in (5) may be calculated. In practice, however, it is typically best to implement these calculations using established software for latent variable modelling. If we have software which can fit the full model using the one-step approach, it can be adapted to produce also the point estimates and their variance matrix for the two-step approach. First, $\tilde{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\Sigma}}_{11}$ are obtained by fitting step 1 latent class model. Second, $\tilde{\boldsymbol{\theta}}_2$ and $\hat{\mathbf{V}}_2$ are obtained by fitting a model which uses the same code as we would use for one-step estimation, except that now the values of $\boldsymbol{\theta}_1$ are fixed at $\tilde{\boldsymbol{\theta}}_1$ rather than estimated. After these steps, the only quantity that remains to be estimated is $\mathcal{I}_{12}$, the cross-parameter block of the information matrix $\mathcal{I}(\boldsymbol{\theta}^*)$. In some applications of PML estimation, this can be an awkward quantity which requires separate calculations. Here, however, it too is easily obtained. This is because software which can fit the one-step model can also evaluate this part of the information matrix. All that we need to do to trick the software into producing the estimate of $\mathcal{I}_{12}$ that we need is to set up estimation of the one-step model with $\tilde{\boldsymbol{\theta}} = (\tilde{\boldsymbol{\theta}}_1, \tilde{\boldsymbol{\theta}}_2)$ as the starting values, get the software to calculate the information matrix with these values (i.e. before carrying out the first iteration of the estimation algorithm), and extract from it the part corresponding to $\mathcal{I}_{12}$. The code included in the supplementary materials for this article shows how this, and the other parts of the two-step estimation can be done in Latent GOLD.

3. Simulation Studies

In this section, we carry out simulation studies to examine the performance of the two-step estimator and to compare it to the existing one-step and three-step estimators. The simulations consider the two specific situations which are discussed in Sect. 2.3 and represented in Fig. 1, i.e. one with models where the latent class is a response variable and one where it is an explanatory variable. The settings of the studies draw on those of previous simulations by Vermunt (2010), Bakk et al. (2013) and Bakk et al. (2014a).

In all of the simulations, there is one latent class variable X with $C = 3$ classes. It is measured by six items $\mathbf{Y} = (Y_1, \ldots, Y_6)$, each with two values which we label 0 and 1. The more likely response is 1 for all six items in class 1, 1 for three items and 0 for three in class 2, and 0 for all items in class 3. The probability of the more likely response is set to the same value π for all classes and items. Higher values of π mean that the association between X and $\mathbf{Y}$ is stronger, separation between the latent classes larger, and precise estimation of the latent class model easier. We use for π the values 0.9, 0.8 and 0.7 and call them the high-, medium-, and low- separation conditions, respectively. Thus, the probabilities of the response 1 are, for example, all 0.9 in class 1 in the high-separation condition and (0.7, 0.7, 0.7, 0.3, 0.3, 0.3) in class 2 in the low-separation condition. The association between X and $\mathbf{Y}$ can be summarized by the entropy-based pseudo-R^2 measure (see e.g. Magidson, 1981): here its value is 0.36, 0.65 and 0.90 in the low-, medium-, and high-separation conditions, respectively. We consider simulations with sample sizes n of 500, 1000, and 2000, resulting in nine sample size-by-class separation simulation settings in each of the two situations we consider.

In the first simulations, the structural model is the multinomial logistic model (3) where the probabilities $p(X = c | Z_p)$ of the latent classes are regressed on a single interval-level covariate Z_p with uniformly distributed integer values 1–5. Class 1 is the reference level for X, and the coefficients for classes 2 and 3 are $\beta_2 = -1$ and $\beta_3 = 1$. The intercepts were set to values yielding equal class sizes when averaged over Z_p. In the second set of simulations, the structural model is a linear regression model with X as the covariate for a continuous response Z_o which is normally distributed with residual variance of 1 (except in one simulation at the end, where violations of this distributional assumption are considered). Omitting the intercept term but including dummy variables for all three latent classes, the regression coefficients $\beta_1 = -1$, $\beta_2 = 1$ and $\beta_3 = 0$ are the expected values of Z_o in classes 1, 2, and 3, respectively.

Mean bias and root-mean-squared error (RMSE) of one regression coefficient (with true value $\beta_3 = 1$) over 500 simulated data sets from a model where a latent class variable X is a response variable for an observed explanatory variable Z_p, under different specifications for the separation between the latent classes and the sample size (please see the main text for more details on the simulation specifications). The results are shown for the proposed two-step estimator and for one-step, naive three-step (with modal assignment) and two bias-adjusted three-step ("BCH" and "ML") estimators.

Class separation	Sample size	MEAN BIAS					RMSE				
		1-step	2-step	3-step			1-step	2-step	3-step		
				Naive	BCH	ML			Naive	BCH	ML
Low	500	.04	$-.24$	$-.59$	$-.25$	$-.27$	.25	.38	.60	.48	.42
	1000	.03	$-.16$	$-.60$	$-.16$	$-.20$	.19	.32	.62	.45	.35
	2000	.01	$-.09$	$-.60$	$-.11$	$-.12$	.12	.22	.61	.36	.24
Medium	500	.01	$-.05$	$-.40$	$-.03$	$-.06$	.17	.20	.41	.30	.22
	1000	.01	$-.01$	$-.37$	.02	$-.02$	.11	.13	.38	.24	.14
	2000	$-.01$	$-.01$	$-.37$	.01	$-.01$	.08	.09	.38	.15	.10
High	500	.02	.01	$-.11$	.02	.01	.13	.13	.17	.16	.14
	1000	.01	.01	$-.12$	.01	.01	.09	.09	.15	.10	.10
	2000	.00	.00	$-.12$	.01	.00	.07	.07	.13	.07	.07

We compare the two-step estimates to ones from the one-step method, the naive three-step method with modal assignment to latent classes in step 2, and the "BCH" and "ML" methods of bias-adjusted three-step estimation. The models were estimated with Latent GOLD version 5.1, with auxiliary calculations done in R (R Core Team, 2016). In each setting, 500 simulated samples were generated. In a small number of samples in the low-separation condition (11 of the 500 when $n = 500$, and 4 when $n = 1000$), one or both of the bias-adjusted three-step methods produced inadmissible estimates (the reasons for this are discussed in Bakk et al., 2013), and these samples are omitted from the results for all estimators. The two-step method produced admissible estimates for all of the samples.

Results of the simulations where X is a response variable are shown in Tables 1 and 2. For simplicity, we report here only results for one of the regression coefficients, which had the true value of $\beta_3 = 1$ (the results for the other coefficient were similar). Table 1 compares the performance of the different estimators of this coefficient in terms of their mean bias and root-mean-squared error (RMSE) over the simulations. We note first that the one-step estimator is essentially unbiased in all the conditions and has the lowest RMSE. The naive three-step estimator is severely biased (and has the highest RMSE), with a bias which decreases with increasing class separation but is unaffected by sample size. The bias-adjusted three-step methods remove this bias, except in cases with low class separation where some of the bias remains. These results are similar to those found by Vermunt (2010).

The two-step estimator is comparable to the bias-adjusted three-step estimators, but consistently slightly better than them. Its smaller RMSE suggests that there is a gain in efficiency from implementing the stepwise idea in this way, avoiding the extra step of three-step estimation. In the medium- and high-separation conditions, the two-step estimator also performs essentially as well as the one-step estimator, suggesting that there is little loss of efficiency from moving from full-information ML estimation to a stepwise approach.

The low-separation condition is the exception to these conclusions. There all of the stepwise estimators have a non-trivial bias and higher RMSE than the one-step estimator (although the two-step estimator is again better than the bias-corrected three-step ones). A similar result was reported for the three-step estimators by Vermunt (2010) and (in simulations where X was a covariate) by Bakk et al. (2013). They concluded that this happens because the first-step estimates

TABLE 2.
Results on estimated sampling variability in the same simulations (and for the same estimated parameter) as in Table 1. Here only the two-step and one-step estimators are compared. The table shows the simulation standard deviation of the point estimates of the parameter (SD), mean of their estimated standard errors (m(SE)) and coverage percentage of 95% confidence intervals (C95). For the two-step estimators, also shown are coverage of 95% confidence intervals when the estimated standard errors include only the uncertainty from the second step of estimation (C95-2), and the average percentage that this standard error contributes to the full standard error (SE%-2).

Class separation	Sample size	1- STEP ESTIMATOR			2- STEP ESTIMATOR				
		SD	m(SE)	C95	SD	m(SE)	C95	SE%-2	C95-2
Low	500	.25	.25	.96	.30	.31	.77	60	.59
	1000	.18	.17	.95	.27	.28	.86	48	.63
	2000	.12	.12	.95	.20	.24	.93	42	.66
Medium	500	.17	.16	.94	.19	.18	.92	80	.85
	1000	.11	.11	.95	.13	.12	.94	86	.90
	2000	.08	.08	.95	.09	.09	.96	88	.92
High	500	.13	.13	.96	.13	.13	.96	98	.96
	1000	.09	.09	.95	.09	.09	.95	99	.94
	2000	.06	.07	.96	.06	.06	.95	99	.95

are biased for the true latent classes when the class separation is low. They also observed that the level of separation in the low condition considered here (where the entropy R^2 is 0.36) would be regarded as very low for practical latent class analysis, i.e. if the observed items $\mathbf{Y}$ were such weak measures of X, they would provide poor support for reliable estimation of associations between the latent class membership and external variables. The one-step estimator performs better because the covariate Z_p in effect serves as an additional indicator of the latent class variable, and indeed one which is arguably stronger than the indicators $\mathbf{Y}$ in the low-separation condition (for example, the standard R^2 for Z_p given X is here 0.48).

In Table 2, we examine the behaviour of the estimated standard errors of the two-step estimators, obtained as explained in Sect. 2.4. We compare them to the one-step estimator (for which the standard errors are obtained from standard ML theory and should behave well), omitting the three-step estimators which are not the focus here (simulation results for their estimated standard errors are reported by Vermunt 2010 and Bakk et al., 2014a).

The first three columns for each estimator in Table 2 show the simulation standard deviation of the estimates of the parameter, the average of their estimated standard errors, and the coverage proportion of 95% confidence intervals calculated using the standard errors. Here the one-step and two-step estimators both behave well in the medium- and high-separation conditions, in that the standard errors are good estimates of the sampling variation and the confidence intervals have correct coverage or very close to it (with 500 simulations, observed coverages between 0.932 and 0.968 are not significantly different from 0.95 at the 5% level). The variability of the estimates is also comparable for the two methods, again indicating that the two-step method is here nearly as efficient as the one-step method. An exception is again the low-separation condition, where the variability of the two-step estimators is higher. Even then their estimated standard errors correctly capture this variability, so the undercoverage of the confidence intervals in the low-separation condition is due to the bias in the two-step point estimator which is shown in Table 1.

The last two columns of Table 2 examine the performance of estimated standard errors of the two-step estimators if they were based only on $\mathbf{V}_2$ in (5), i.e. if we ignored the contribution from the uncertainty from the first step of estimation which is captured by $\mathbf{V}_1$. The "C95-2" column of the table shows the coverage of 95% confidence intervals if we do this, and "SE%-2" shows the percentage that step 2 only standard errors contributes to the full standard errors (this

TABLE 3.

Simulation results for point estimates of one regression coefficient (with true value $\beta_2 = 1$) over 500 simulated data sets from a model where a latent class variable X is an explanatory variable for an observed response variable Z_o. The table shows the same quantities as Table 1.

Class separation	Sample size	MEAN BIAS					RMSE				
		1-step	2-step	3-step			1-step	2-step	3-step		
				Naive	BCH	ML			Naive	BCH	ML
Low	500	.01	− .32	− .66	− .38	− .36	.19	.44	.70	.50	.47
	1000	.01	− .16	− .60	− .22	− .20	.12	.30	.63	.37	.33
	2000	.00	− .08	− .58	− .12	− .11	.08	.21	.59	.27	.23
Medium	500	.00	− .03	− .31	− .03	− .03	.11	.13	.33	.16	.13
	1000	.00	− .01	− .29	.01	− .01	.08	.09	.30	.12	.09
	2000	.00	.01	− .29	.00	.00	.05	.07	.29	.09	.07
High	500	.00	.00	− .08	.00	.00	.08	.09	.12	.09	.09
	1000	.01	.01	− .07	.01	.01	.06	.06	.10	.07	.06
	2000	.00	.00	− .08	.00	.00	.04	.04	.09	.05	.04

TABLE 4.

Results on estimated sampling variability in the same simulations (and for the same estimated parameter) as in Table 3. The table shows the same quantities as Table 2.

Class separation	Sample size	1- STEP ESTIMATOR			2- STEP ESTIMATOR				
		SD	m(SE)	C95	SD	m(SE)	C95	SE%-2	C95-2
Low	500	.19	.16	.89	.30	.31	.77	46	.43
	1000	.12	.11	.94	.25	.27	.88	40	.47
	2000	.08	.08	.94	.19	.20	.88	32	.52
Medium	500	.11	.11	.95	.12	.12	.93	76	.88
	1000	.08	.07	.95	.09	.08	.94	74	.85
	2000	.05	.05	.95	.07	.06	.95	73	.85
High	500	.09	.08	.94	.09	.09	.95	99	.94
	1000	.06	.06	.95	.06	.06	.96	99	.95
	2000	.04	.04	.95	.04	.05	.95	99	.94

is calculated by comparing the simulation averages of these two kinds of standard errors). It can be seen that in the low-separation conditions around half of the uncertainty actually arises from step 1 estimates, and ignoring this results in severe underestimation of the true uncertainty and very poor coverage of the confidence intervals. Even in the more sensible medium-separation condition the contribution from step 1 uncertainty is over 10% and the coverage is non-trivially reduced, and it is only in the high-separation condition that we could safely treat step 1 estimates as known. These results suggest that there is a clear benefit from using standard errors calculated from the full variance matrix (5) derived from pseudo-ML theory.

Tables 3 and 4 show the same statistics for the simulations where the latent class X is an explanatory variable for a continuous response Z_o. Here we again focus on just one parameter in this model, with true value $\beta_2 = 1$. The results of these simulations are very similar to the ones where X was the response variable (and for the one-step and three-step estimators they are also similar to the results in Bakk et al., 2013). The two-step estimator again performs a little better

Simulation results for point estimates of three regression coefficients over 500 simulated data sets from a model where a latent class variable X is an explanatory variable for an observed response variable Z_o. The simulation settings are the same as in the case with medium separation and sample size 1000 in Table 3, except that here the residual distribution of Z_o is not normal and homoscedastic but skewed, heteroscedastic or bimodal (see the main text for how these are specified). The results are shown for all three regression coefficients (βs) in the model for Z_o (which have true values of -1, 1, and $+1$).

Residual distribution	Value of β	MEAN BIAS					RMSE				
		1-step	2-step	3-step			1-step	2-step	3-step		
				Naive	BCH	ML			Naive	BCH	ML
Skewed	-1	$-.05$	$-.01$	.23	.01	$-.03$	.09	.09	.25	.11	.10
	0	$-.05$	$-.02$	.07	.00	$-.08$	.09	.09	.09	.08	.12
	$+1$	.03	.02	$-.29$	.00	.06	.08	.10	.30	.12	.11
Bimodal	-1	.02	.03	.24	.01	.04	.07	.09	.25	.11	.10
	0	.01	.02	.07	.00	.01	.07	.07	.09	.08	.08
	$+1$	$-.02$	$-.03$	$-.29$	.00	$-.04$	.08	.10	.30	.12	.11
Heteroscedastic	-1	.12	.12	.23	.00	.13	.14	.15	.24	.11	.16
	0	$-.45$	$-.03$	.06	.01	$-.06$	.83	.17	.14	.15	.22
	$+1$	$-.03$	$-.10$	$-.29$	.00	$-.09$	.14	.15	.30	.13	.15

than the three-step estimators and, except in situations with low class separation, essentially as well as the one-step estimator.

It is also of interest to consider to what extent the different estimators may be sensitive to misspecifications of different parts of the models. In a further simulation, for which the point estimates are shown in Table 5, we examine this with respect to violations of the assumptions about the distribution of a continuous outcome Z_o. Here the settings are the same as in the medium-separation condition with the sample size of 1000 in Table 3, except that the true distribution of the residuals in the model for Z_o given X differs in one of three ways from the homoscedastic normal distribution which is assumed by the estimators. In the first case, it is a mixture of the normal distributions $N(-0.5, 0.15)$ and $N(0.75, 1.3375)$, with weights 0.6 and 0.4, respectively; this has variance 1 but is positively skewed, with an index of skewness of 1.16 (roughly the same as that of the χ_6^2 distribution). In the second, it is a mixture of $N(0.9, 0.19)$ and $N(-0.9, 0.19)$ with equal weights; this is symmetric with variance 1, but is very clearly bimodal. In the third case, the residual distribution is normal but heteroscedastic, in that its variance is 1 in two of the latent classes but 5 in one of them.

We may anticipate that the BCH and naive three-step methods should be robust in this respect, because in step 3 they use standard linear regression (weighted or unweighted) for Z_o given assigned values of X, which does not rely on parametric assumptions about the residual distribution. Table 5 shows that this is indeed the case, and for these estimators the results are essentially the same as in Table 3. In contrast, the one- and two-step methods and the 3-step ML method each use in their last step a log likelihood which involves a distribution for Z_o. Their estimates may then become biased when the fitted model tries to reconcile the assumed homoscedastic normal distribution of Z_o with the observed data, and for one-step estimates this bias may be further increased because the method allows the latent classes themselves to be affected by the observed distribution of Z_o. Here these effects are, however, small in the cases with skewed or bimodal distributions. For them, the two-step and one-step estimates remain comparable in RMSE, and somewhat better than the adjusted three-step estimates.

In the case of a heteroscedastic residual distribution, the lowest RMSEs are achieved by the BCH estimates. All the other estimates have some bias, which affects different class-specific

mean parameters differently (and for one of them has a very large bias for the one-step estimates). This case suggests that the one- and two-step methods may be most sensitive when the violations of the distributional assumptions vary by latent class; for one-step estimates, results for such situations are also reported by Bakk and Vermunt (2016) and Asparouhov and Muthén (2015), from simulations which use class-specific and severely bimodal residual distributions. It should be noted, however, that these model violations are of a kind which should in practice not go unobserved by the data analyst, but would be easily detectable even from a preliminary analysis with naive three-step estimation.

4. Empirical Examples

4.1. Latent Class as a Response Variable: Tolerance Towards Nonconformity

In this first applied example, we consider a latent class analysis of items which measure intolerance towards different groups of others. The substantive research question is whether different levels and patterns of intolerance are associated with individuals' education and birth cohort. We use data from the 1976 and 1977 US General Social Surveys (GSS) which was first analysed by McCutcheon (1985) using the naive three-step method with modal assignment of latent classes to individuals. Bakk et al. (2014a) re-analysed the data using the one-step and bias-corrected three-step methods, thus showing how McCutcheon's original estimates are affected when the misclassification from the second-step class allocation is taken into account. We examine how two-step estimates compare with these previously proposed approaches in this example.

For the definitions of variables and for the first-step latent class modelling, we follow the choices made by the previous authors (data and code for the analysis of Bakk et al., 2014a is given in Bakk, Oberski, & Vermunt, 2014b, and for our analysis in the supplementary materials for this article). The original survey measured a respondent's tolerance of communists, atheists, homosexuals, militarists and racists, using three items for each of these groups. The items asked if the respondent thought that members of a group should be allowed to make speeches in favour of their views, teach in a college, and have books written by them included in a public library (the wordings of the questions are given in McCutcheon, 1985). Thus, "tolerance" here essentially means willingness to grant members of a group public space and freedom to disseminate their views. McCutcheon recoded the data into five dichotomous items, one for each group, by coding the attitude towards a group as tolerant if the respondent gave a tolerant answer to all three items for that group, and intolerant otherwise.

The first-step latent class analysis is carried out on a sample of 2689 respondents who had an observed value for all five items. This complete-case analysis was used to match that of McCutcheon (1985). It is not essential, however, and all of the estimators can also accommodate observations with missing values in some of the items (we will do that in our second example in Sect. 4.2). There were further 21 respondents who are excluded from estimation of the structural model because they had missing values for the covariates.

We use the same four-class latent class model for the tolerance items which was also employed by the previous authors. Its estimated parameters are shown in Table 6. The upper part of the table gives the estimated parameters of the measurement model, that is the probabilities $\pi_{kc1} = P(Y_{ik} = 1|X_i = c)$ that a respondent i who belongs to latent class c gives a response which is coded as tolerant of group k. Using the labels introduced by McCutcheon, the class in the first column is called "Tolerant" since respondents in this class have a high probability of being tolerant of all five groups. The "Intolerant of Right" class is intolerant of groups such as racists and militarists and the "Intolerant of Left" class particularly intolerant of communists, while the "Intolerant" has a low probability of a tolerant response for all five groups. The entropy-based pseudo-R^2 measure is here 0.72, placing the separation of these classes between the medium- and high-separation

Estimated probability parameters for the four-class latent class model for survey items on tolerance towards different groups. Here 'Class size' refers to probabilities $\rho_c = P(X = c)$ of the latent classes, and the other numbers in the table are the probabilities of giving a tolerant response to an item given the membership of a latent class.

	Latent class			
	"Tolerant"	"Intolerant of Right"	"Intolerant of Left"	"Intolerant"
Probability of tolerance for...				
Atheists	.98	.41	.61	.03
Communists	.95	.59	.27	.02
Militarists	.92	.34	.38	.05
Racists	.90	.02	.81	.08
Homosexuals	.96	.72	.56	.13
Class size	.23	.11	.10	.56

conditions in our simulations in Sect. 3. The last row of the table gives the estimated probabilities ρ of the latent classes; these show that the Intolerant class is the largest, with a probability of 0.56.

The structural models are multinomial logistic models (3) for these latent classes given a respondent's education and birth cohort (which in these cross-sectional data is indistinguishable from age). Educational attainment was coded into three categories, based on years of formal education completed: less than 12 ("Grade school"), 12 ("High school"), or more than 12 years ("College"). Birth cohort was coded by McCutcheon into four categories: those born after 1951 (and thus aged 17–23 in 1976), in 1934–51 (24–42), 1915–33 (43–61), or before 1915 (62 or older). Here we treat this variable as continuous for simplicity of presentation, with values 1–4, respectively.

The estimates of the structural model are shown in Table 7, in the form of estimated coefficients for being in the other three classes relative to the Tolerant class. Consider first the estimates from the stepwise approaches, which are here all fairly similar to each other. The overall Wald tests show that both education and birth cohort have clearly significant associations with membership of the different tolerance classes. People from the older cohorts are more likely to be in the Intolerant of Left and (especially) the Intolerant classes, but there is no significant cohort effect on being in the Intolerant of Right rather than Tolerant class. Having college education rather than either of the two lower levels of education is very strongly associated with lower probabilities of all of the three intolerant classes, and the same is true for high school vs. grade school education in the comparison of Intolerant and Intolerant of Left against the Tolerant class (the latter contrast is significant only for the two-step and naive three-step estimates).

The one-step estimates in Table 7 are rather more different from all the stepwise estimates. This difference arises from a deeper discrepancy than just that of different estimates for the same parameters. Here the parameters are in fact not the same, because the one-step estimates are effectively coefficients for a different response variable. This point is demonstrated in Table 8. It shows the estimated measurement probabilities and marginal class sizes of the latent class model from one-step estimation with different choices of the covariates in the structural model. The first column for each class shows the results when no covariates are included, so it is the same as the model in Table 6 (with the classes there numbered here 1–4 in the same order). We refer to this pattern and interpretation of the classes as "pattern A". The estimates from one-step estimation follow this pattern also if the structural model includes only the birth cohort, or the cohort plus education included as years completed rather than in the grouped form. In other words, in these cases the one-step estimates of the measurement probabilities of the latent classes are sufficiently

Estimated coefficients (with estimated standard errors in parentheses) of the multinomial logistic structural model in the example in Sect. 4.1, for the latent class of tolerance towards different groups given a respondent's education and birth cohort.

Latent class (vs. Tolerant)	Covariate	Estimator				
		2-step	1-step[†]	3-step		
				Naive	BCH	ML
Intolerant of Right	Education					
	High school	0.12 (0.36)	0.38 (0.44)	-0.12 (0.21)	-0.04 (0.31)	-0.08 (0.30)
	College	-1.07 (0.39)	-1.12 (0.46)	-1.23 (0.20)	-1.31 (0.31)	-1.29 (0.29)
	Cohort	0.07 (0.13)	0.87 (0.10)	0.12 (0.09)	-0.01 (0.14)	-0.04 (0.15)
Intolerant of Left	Education					
	High school	-0.65 (0.29)	-0.35 (0.23)	-0.45 (0.21)	-0.53 (0.31)	-0.41 (0.31)
	College	-2.06 (0.30)	-1.69 (0.23)	-1.55 (0.22)	-1.80 (0.32)	-1.75 (0.31)
	Cohort	0.36 (0.16)	0.23 (0.09)	0.42 (0.10)	0.44 (0.14)	0.42 (0.14)
Intolerant	Education					
	High school	-0.77 (0.19)	-1.83 (0.38)	-0.61 (0.15)	-0.76 (0.19)	-0.72 (0.19)
	College	-2.33 (0.19)	-3.75 (0.49)	-1.94 (0.14)	-2.25 (0.18)	-2.22 (0.17)
	Cohort	0.97 (0.08)	1.14 (0.13)	0.82 (0.06)	0.96 (0.08)	0.96 (0.08)
p values of overall Wald tests of the covariates						
	Education ($df = 6$)	$< .001$	$< .001$	$< .001$	$< .001$	$< .001$
	Cohort ($df = 3$)	$< .001$	$< .001$	$< .001$	$< .001$	$< .001$

† *Note:* The latent classes implied by the 1-step estimates are not really the same as for the other methods. Please see Table 8 for the measurement model of this model, and the discussion in the text.

similar from one model to the next so that the interpretation (and labelling) of the classes remains unchanged, even though the exact values of these probabilities still change between models.

In other models, however, the estimated measurement model changes so much that the latent classes themselves change. We refer to these cases in Table 8 as "pattern B" (nearest matches from the two patterns are shown under the same number of class in the table). In this pattern, the Tolerant class maintains its interpretation and estimated size, but the other three classes are re-arranged so that we end up with two classes (numbers 2 and 4) with slightly different patterns of low tolerance and one class (3) with a probability of a tolerant response around 50% for all the groups. This pattern emerges when the structural model includes education alone in either years completed or in the grouped form. It also appears when the covariates are cohort and the grouped education, which was the model we considered in Table 7. The one-step model there is thus a model for latent classes of pattern B (with the measurement probabilities shown in the second column for each class in Table 8), whereas all the stepwise models are for classes of pattern A.

This example illustrates the inherent property of one-step estimation that every change in the structural model will also change the measurement model. Sometimes these changes are small, such as those between the different versions of pattern A in Table 8, but sometimes they are so large, such as the jumps between patterns A and B, that they effectively change the meaning of the latent class variable. There is no reason even to expect that the possible patterns would be limited to two as here, so in analyses with a larger number of covariates still more patterns could appear. In practical analysis, it could happen that the analyst failed to notice these changes and hence to realize that comparisons between some structural models were effectively meaningless. Even if the analyst did pay attention to this feature, there is nothing they could really do about it within one-step estimation. This is because the method provides no entirely coherent way of

TABLE 8.

Estimated class sizes and probabilities of giving a tolerant response towards different groups, in four-class latent class models estimated as part of one-step estimation with different covariates for the latent class variable included (+) or not included (−) in the model. The first model, with no covariates, is the model obtained from step 1 of stepwise approaches and shown also in Table 6. The second model, with cohort and grouped education variable, is the measurement model from the 1-step estimation in Table 7. Two broad patterns of the latent classes appear here, labelled in the table as patterns A and B.

Covariates												
Cohort	−	+	−	+	−	+	−	+	−	+	−	+
Education (grouped)	−	+	+	−	−	−	−	+	+	−	−	−
Education (years)	−	−	−	−	+	+	−	−	−	−	+	+
Class pattern	A	B	B	A	B	A	A	B	B	A	B	A
	Class 1						Class 2					
Atheists	.98	1.00	1.00	.99	1.00	.99	.41	.01	.02	.40	.02	.36
Communists	.95	.94	.95	.95	.95	.95	.59	.10	.10	.46	.11	.46
Militarists	.92	.94	.94	.92	.95	.93	.34	.04	.04	.30	.03	.28
Racists	.90	.91	.91	.89	.91	.89	.02	.00	.00	.00	.00	.00
Homosexuals	.96	.96	.96	.96	.96	.95	.72	.22	.23	.60	.25	.62
Class size[†]	.23	.21	.21	.23	.21	.22	.11	.28	.28	.16	.25	.16
	Class 3						Class 4					
Atheists	.61	.55	.56	.62	.57	.63	.03	.02	.03	.02	.04	.02
Communists	.27	.42	.44	.30	.45	.32	.02	.00	.01	.03	.02	.03
Militarists	.38	.37	.38	.38	.39	.39	.05	.07	.07	.05	.08	.05
Racists	.81	.42	.45	1.00	.45	.96	.08	.16	.16	.09	.15	.09
Homosexuals	.56	.63	.66	.58	.68	.60	.13	.05	.05	.13	.05	.12
Class size[†]	.10	.23	.22	.09	.22	.09	.56	.27	.29	.53	.32	.52

† Obtained by averaging conditional class probabilities over the sample distribution of the covariates.

forcing the measurement model to remain the same. In contrast, all stepwise methods achieve this by definition, because their key feature is that the measurement model is fixed before any structural models are estimated.

4.2. Latent Class as an Explanatory Variable: Psychological Contract Types and Job Insecurity

Our second example draws on the Dutch and Belgian samples of the Psychological Contracts across Employment Situations project (PSYCONES, 2006). These data were used by Bakk et al. (2013) to compare the one-step and bias-adjusted three-step approaches, and we follow their choices for the models and variables. The goal is to examine the association between an individual's perceived job insecurity and their perception of their own and their employee's obligations in their current employment (the "psychological contract"). Job insecurity is measured on a scale used by the PSYCONES project (originally from De Witte, 2000), treated as a continuous variable. Psychological contract types are measured by eight dichotomous survey items. Four of them refer to perceived obligations (promises given) by the employer and four to obligations by the employee, and in each group of four, two items refer to relational and two to transactional obligations. The labels in Table 9 give an idea of the items' content, and their full wordings are given by De Cuyper, Rigotti, Witte, and Mohr (2008) who also analysed these items (for a different sample) with latent class analysis. We derive a classification of psychological contract types from a latent class model and use it as a covariate in the structural model which is a linear regression model for perceived job insecurity.

TABLE 9.
Estimated probability parameters for the four-class latent class model for survey items on perceived psychological contract
of employment. Here 'Class size' refers to probabilities of the latent classes, and the other numbers in the table are the
conditional probabilities (given each latent class) of believing that a particular type of obligation has been given by the
employer or the employee.

	Latent class			
	"Mutual High"	"Over-obligation"	"Under-obligation"	"Mutual Low"
Employer's obligations				
Secure job	.90	.37	.87	.21
Advancement	.89	.30	.84	.17
Good pay	.87	.29	.75	.27
Safe work environment	.98	.55	.73	.29
Employee's obligations				
Loyalty	.96	.73	.37	.08
Volunteering	.96	.83	.37	.17
Being on time	.98	.96	.38	.18
Good performance	1.00	.97	.77	.28
Class size	.53	.29	.09	.09

There are 1431 respondents who answered at least one of the eight items, and all of them
are used for the first-step latent class modelling. In general, all of the methods considered here
can accommodate units of analysis which have missing data in some of the items. For estimation
steps which employ a log likelihood of some kind (such as one-step estimation and both steps of
two-step estimation), this is done by defining it in such a way that all observed variables contribute
to the log likelihood for each unit, and for the second step of the three-step methods it is achieved
by calculating the conditional probability of latent classes given all the observed items for each
unit. Four respondents for whom the measure of job insecurity was not recorded are omitted when
the structural model is estimated.

Step 1 model is a four-class latent class model, for which the parameter estimates are given in
Table 9. The first class, which consists of an estimated 52% of the individuals and is labelled the
class of "Mutual High" obligations, is characterized by a high probability of thinking that both the
employer and the employee have given obligations to each other. The "Under-obligation" class
(10%) is likely to perceive that obligations were given by the employer but not the employee, the
opposite is the case in the "Over-obligation" class (29%), and the "Mutual Low" class (9%) have a
low probability of perceiving that any obligations have been given or received. The entropy-based
R^2 for this model is 0.71, which is again between the medium- and high-separation conditions in
our simulation studies.

Estimated coefficients of the structural model are shown in Table 10. Here the naive three-step
estimates are the most different, in that they are closer to zero than are the other estimates. The
rest of the estimates are similar, and the one-step ones are now also comparable to the rest because
their estimated measurement model (not shown) implies essentially the same latent classes as the
first-step estimates used by the stepwise approaches. The estimated coefficients show that the
expected level of perceived job insecurity is similar (and not significantly different) in the Mutual
High and Under-obligation classes and significantly higher in the Over-obligation and Mutual
Low classes (which do not differ significantly from each other). In other words, employees tend
to feel more secure in their job whenever they perceive that the employer has made a commitment
to them, whereas an employee's perception of their own level of commitment has no association
with their insecurity.

TABLE 10.

Estimated coefficients for a linear regression model for perceived job insecurity given latent classes of types of psychological contract. Here the class of "Mutual High" obligation is the reference category. The p value in the last column is for the Wald test (with 3 degrees of freedom) that the coefficients for the other three classes are all zero.

Estimator	Coefficient (with SE) of latent class (vs. Mutual High)						p value[†]
	Over-obligation		Under-obligation		Mutual Low		
2-step	0.51	(0.08)	-0.11	(0.12)	0.45	(0.11)	$< .001$
1-step	0.55	(0.08)	-0.16	(0.12)	0.48	(0.11)	$< .001$
3-step							
Naive	0.39	(0.06)	-0.06	(0.09)	0.37	(0.10)	$< .001$
BCH	0.49	(0.08)	-0.11	(0.12)	0.43	(0.11)	$< .001$
ML	0.51	(0.08)	-0.11	(0.10)	0.43	(0.12)	$< .001$

5. Discussion

The stepwise approaches that we have explored in this article are motivated by the principle that definitions of variables should be separated from the analyses that use them. This is natural and goes unmentioned in most applications where variables are treated as directly observable, where they are routinely defined and measured first and only then used in analysis. Things are not so straightforward in modelling with latent variables, where these variables are defined by their estimated measurement models. One-step methods of modelling do not follow the stepwise principle but estimate simultaneously both the measurement models and the structural models between variables. As a result, the interpretation of the latent variables may change from one model to the next, possibly dramatically so. Stepwise methods of modelling avoid this problem by fixing the measurement model at its value estimated from their first step. In naive three-step estimation, this incurs a bias because the derived variables used in its third step are erroneous measures of the variables defined in the first step. This bias is removed by the bias-adjusted three-step and the two-step methods. In this article, we have argued that the two-step method that we have proposed is the more straightforward of them and has somewhat better statistical properties.

Other properties of the two-step method remain to be studied further. These include, for example, its robustness to violations of assumptions in different parts of the joint model. In Sect. 3, we examined this briefly with respect to distributional assumptions about a continuous response variable, with results which suggested that two-step estimates are fairly robust in this respect, somewhat more so than one-step estimates but less so than some three-step estimates. The conclusions may be different for other parts of the structural and measurement models. A particularly important question of this kind is the assumption that the measurement model depends only on the latent class X but not on other variables $\mathbf{Z}$. This is the assumption of measurement equivalence (absence of differential item functioning), which may be violated in many applications. Here observing that one-step estimates change when variables in the structural model are changed may itself be a sign that the measurement model is misspecifed in this respect. Questions of interest about non-equivalence of measurement are not limited to the sensitivity of estimates if it is wrongly ignored, but include also how two-step estimation could be used to detect non-equivalence and to estimate models which allow for it. This is an important topic for future research on the two-step approach.

We have focused on latent class analysis, but both the methods and the principles that we have described apply also more generally. They could be extended to models with other kinds of latent variables, such as linear structural equation models (SEMs) where both the latent variables and their measures are treated as continuous. In this context, the one-step method (conventional

SEMs) and the naive three-step method (using factor scores as derived variables) are routinely used, while other stepwise methods are not fully developed. There too the one-step approach has the property that the measurement models of the latent factors do not remain fixed, although it could be that the consequences of this are less dramatic than they can be for the categorical latent variables in latent class analysis. Two-step estimation can be defined and implemented for models with continuous latent variables in the same way as described in this article for latent classes, in effect by making the appropriate changes to the distributions defined in our Sect. 2. The behaviour of the two-step approach in this context remains to be investigated.

Appendix: Score Functions and Information Matrices for Latent Class Models

Consider first in general terms a model which involves a latent class variable X with a total of C latent classes (here X may also represent all combinations of the classes of several latent class variables). Suppose that the model depends on parameters $\boldsymbol{\theta}$ (in this Appendix, we take all vectors to be column vectors, the opposite of the practice in Sect. 2 where they were row vectors for simplicity of notation). The log likelihood contribution for a single unit i is then $l_i = \log L_i = \log \sum_{c=1}^{C} L_{ic}$, where $L_{ic} = \exp(l_{ic})$ is the term in L_i which refers to latent class $c = 1, \ldots, C$. The contribution of unit i to the score function is then $\mathbf{u}_i = \partial l_i / \partial \boldsymbol{\theta} = \mathbf{h}_i / L_i$ where $\mathbf{h}_i = \partial L_i / \partial \boldsymbol{\theta} = \sum_c L_{ic} \mathbf{u}_{ic}$ and $\mathbf{u}_{ic} = \partial l_{ic} / \partial \boldsymbol{\theta}$, and the contribution to the observed information matrix is

$$\mathbf{J}_i = -\frac{\partial^2 l_i}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = -\frac{1}{L_i^2} \left\{ L_i \left[\sum_c L_{ic} \left(\mathbf{u}_{ic} \mathbf{u}_{ic}' - \mathbf{J}_{ic} \right) \right] - \mathbf{h}_i \mathbf{h}_i' \right\},$$

where $\mathbf{J}_{ic} = -\partial^2 l_{ic} / (\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}')$.

Suppose that observations for different units $i = 1, \ldots, n$ are independent. Point estimation of $\boldsymbol{\theta}$ is easiest with the EM algorithm (Dempster, Laird, & Rubin, 1977). For this, let l_i^* be the same expression as l_{ic} but now regarded as a function of c. At the E step of the $(t+1)$th iteration of EM, we calculate $Q^{(t+1)} = \sum_i \mathrm{E}[l_i^* | \mathbf{D}, \boldsymbol{\theta}^{(t)}] = \sum_i \left(\sum_c \pi_{ic}^{(t)} l_{ic} \right)$ where $\pi_{ic}^{(t)} = p(X_i = c | \mathbf{D}, \boldsymbol{\theta}^{(t)})$, $\mathbf{D}$ denotes all the observed data, and $\boldsymbol{\theta}^{(t)}$ is the estimate of $\boldsymbol{\theta}$ from the tth iteration. At the M step, $Q^{(t+1)}$ is maximized with respect to $\boldsymbol{\theta}$ to produce an updated estimate $\boldsymbol{\theta}^{(t+1)}$. This is relatively straightforward because $\partial Q^{(t+1)} / \partial \boldsymbol{\theta} = \sum_i \sum_c \left(\pi_{ic}^{(t)} \mathbf{u}_{ic} \right)$ and $-\partial^2 Q^{(t+1)} / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}' = \sum_i \sum_c \left(\pi_{ic}^{(t)} \mathbf{J}_{ic} \right)$, i.e. these are the score function and observed information matrix for a model where X is known, fitted to pseudo-data of $n \times C$ observations with fractional weights $\pi_{ic}^{(t)}$.

The information matrix $\mathcal{I}$ for the model can be estimated by $n^{-1} \sum_i \mathbf{J}_i$ or $n^{-1} \sum_i \mathbf{u}_i \mathbf{u}_i'$. Together with $\mathbf{u}_i$, these could also be used to implement other estimation algorithms than EM. When evaluated at the final estimate of $\boldsymbol{\theta}$, they give estimates of the $\mathcal{I}_{22}$ and $\mathcal{I}_{12}$ which are needed for the two-step variance matrix (5). An estimate of the Σ_{11} that is also needed there is obtained similarly from the estimated information matrix for step 1 latent class model.

What then remains to be done for any specific model is to evaluate l_{ic}, $\mathbf{u}_{ic}$ and (if used) $\mathbf{J}_{ic}$ for it. As an example, consider the model with covariates $\mathbf{Z}_p$, one latent class variable X, and a response variable Z_o which is considered in Sect. 2.1. The L_i for it is given by (2) with $p(\mathbf{Z}_{pi})$ omitted. Then $l_{ic} = \log p(X_i = c | \mathbf{Z}_{pi}) + \log p(Z_{oi} | \mathbf{Z}_{pi}, X_i = c) + \sum_k \log p(Y_{ik} | X_i = c) \equiv l_{ic}^{(x)} + l_{ic}^{(z)} + \sum_k l_{ic}^{(yk)}$. If the parameters for the different components of the model are distinct, which should be the case for most sensible models, we only need to consider the separate derivatives of the terms in this sum. Suppose that X_i given $\mathbf{Z}_{pi}$ is given by the multinomial logistic model (3),

writing it now as $p(X_i = c | \mathbf{Z}_{pi}) \equiv \pi_{ic}^{(x)} = \exp(\boldsymbol{\alpha}_c' \mathbf{Z}_i^*) / \sum_{s=1}^{C} \exp(\boldsymbol{\alpha}_s' \mathbf{Z}_i^*)$ for $c = 1, \ldots, C$, with $\boldsymbol{\alpha}_1 = \mathbf{0}, \boldsymbol{\alpha}_c = (\beta_{0c}, \boldsymbol{\beta}_c')'$ for $c \neq 1$, and $\mathbf{Z}_i^* = (1, \mathbf{Z}_{pi}')'$. Then $l_{ic}^{(x)} = \log \pi_{ic}^{(x)}, \partial l_{ic}^{(x)} / \partial \boldsymbol{\alpha}_r = (I(r = c) - \pi_{ir}^{(x)}) \mathbf{Z}_i^*$, and $\partial^2 l_{ic}^{(x)} / \partial \boldsymbol{\alpha}_r \partial \boldsymbol{\alpha}_s' = -(I(s = r) \pi_{ir}^{(x)} - \pi_{ir}^{(x)} \pi_{is}^{(x)}) \mathbf{Z}_i^* \mathbf{Z}_i^{*'}$ for $r, s = 2, \ldots, C$. The measurement models for the items Y_{ik} can also be formulated as multinomial logistic models, by writing them as $p(Y_{ik} = r | X_i = c) \equiv \pi_{icr}^{(yk)} = \exp(\boldsymbol{\gamma}_{kr}' \mathbf{X}_{ic}^*) / \sum_{s=1}^{R_k} \exp(\boldsymbol{\gamma}_{ks}' \mathbf{X}_{ic}^*)$ for $r = 1, \ldots, R_k$, with $\mathbf{X}_{ic}^* = (I(c = 1), \ldots, I(c = C))'$ and $\boldsymbol{\gamma}_{k1} = \mathbf{0}$. Then $l_{ic}^{(yk)} = \sum_{r=1}^{R_k} I(Y_{ik} = r) \log \pi_{icr}^{(yk)}$, and if the parameters for different items k are also distinct, the terms in $\sum_k l_{ic}^{(yk)}$ can be differentiated separately. Their derivatives are $\partial l_{ic}^{(yk)} / \partial \boldsymbol{\gamma}_r = (I(Y_{ik} = r) - \pi_{icr}^{(yk)}) \mathbf{X}_{ic}^*$ and $\partial^2 l_{ic}^{(yk)} / \partial \boldsymbol{\gamma}_r \partial \boldsymbol{\gamma}_s' = -(I(s = r) \pi_{icr}^{(yk)} - \pi_{icr}^{(yk)} \pi_{ics}^{(yk)}) \mathbf{X}_{ic}^* \mathbf{X}_{ic}^{*'}$ for $r, s = 2, \ldots, R_k$. For $l_{ic}^{(z)}$, suppose, for example, that Z_{oi} is normally distributed with mean $\mu_i = \boldsymbol{\delta}' \mathbf{Z}_i^{**}$ and variance τ^{-1}, where $\mathbf{Z}_i^{**} = (\mathbf{X}_{ic}', \mathbf{Z}_{pi}')'$. Defining $e_i = Z_{oi} - \mu_i$, then $l_{ic}^{(z)} = (\log \tau - \tau e_i^2)/2, \partial l_{ic}^{(z)} / \partial \boldsymbol{\delta} = \tau e_i \mathbf{Z}_i^{**}, \partial l_{ic}^{(z)} / \partial \tau = (1/\tau - e_i^2)/2, \partial^2 l_{ic}^{(z)} / \partial \boldsymbol{\delta} \partial \boldsymbol{\delta}' = -\tau (\mathbf{Z}_i^{**})(\mathbf{Z}_i^{**})', \partial^2 l_{ic}^{(z)} / \partial^2 \tau = -1/(2\tau^2)$, and $\partial^2 l_{ic}^{(z)} / \partial \boldsymbol{\delta} \partial \tau = e_i \mathbf{Z}_i^{**}$. The formulas for the situations considered in our simulations and examples are obtained from these results by setting $\mathbf{Z}_i^* = 1$ for the case with no Z_{pi}, and omitting $l_{ic}^{(z)}$ for the case with no Z_{oi}. Finally, doing both of these things gives the formulas for the basic latent class model which is estimated in step 1 of the two-step method.

References

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, *103*, 411–423.

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

Asparouhov, T., & Muthén, B. (2015). *Auxiliary variables in mixture modeling: Using the BCH method in Mplus to estimate a distal outcome model and an arbitrary secondary model* (Mplus Web Notes No. 21).

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

Bakk, Z., Oberski, D. L., & Vermunt, J. K. (2014b). *Replication data for: Relating latent class assignments to external variables: Standard errors for correct inference.* Harvard Dataverse. Retrieved from https://doi.org/10.7910/DVN/24497.

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*, *43*, 272–311.

Bakk, Z., & Vermunt, J. K. (2016). Robustness of stepwise latent class modeling with continuous distal outcomes. *Structural Equation Modeling*, *23*, 20–31.

Bandeen-Roche, K., Miglioretti, D. L., Zeger, S. L., & Rathouz, P. J. (1997). Latent variable regression for multiple discrete outcomes. *Journal of the American Statistical Association*, *92*, 1375–1386.

Bartolucci, F. , Montanari, G. E., & Pandolfi, S. (2014). A comparison of some estimation methods for latent Markov models with covariates. In *Proceedings of COMPSTAT 2014—21st International Conference on Computational Statistics* (pp. 531–538).

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

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Bollen, K. A. (1996). An alternative two stage least squares (2SLS) estimator for latent variable equations. *Psychometrika*, *61*, 109–121.

Burt, R. S. (1973). Confirmatory factor-analytic structures and the theory construction process. *Sociological Methods & Research*, *2*, 131–190.

Burt, R. S. (1976). Interpretational confounding of unobserved variables in structural equation models. *Sociological Methods & Research*, *5*, 3–52.

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and applications*. Cambridge: Cambridge University Press.

Chan, T. W., & Goldthorpe, J. H. (2007). European social stratification and cultural consumption: Music in England. *Sociological Review*, *23*, 11–19.

Clogg, C. C. (1981). New developments in latent structure analysis. In D. J. Jackson & E. F. Borgotta (Eds.), *Factor analysis and measurement in sociological research*. London: Sage.

Croon, M. (2002). Using predicted latent scores in general latent structure models. In G. A. Marcoulides & I. Moustaki (Eds.), *Latent variable and latent structure models* (pp. 195–223). Mahwah, NJ: Lawrence Erlbaum.

Dayton, C. M., & Macready, G. B. (1988). Concomitant-variable latent class models. *Journal of the American Statistical Association*, *83*, 173–178.

De Cuyper, N., Rigotti, T., Witte, H. D., & Mohr, G. (2008). Balancing psychological contracts: Validation of a typology. *International Journal of Human Resource Management*, *19*, 543–561.

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

Devlieger, I., Mayer, A., & Rosseel, Y. (2016). Hypothesis testing using factor score regression: A comparison of four methods. *Educational and Psychological Measurement*, *76*, 741–770.

De Witte, H. (2000). Arbeidsethos en jobonzekerheid: Meting en gevolgen voor welzijn, tevredenheid en inzet op het werk [Work ethic and job insecurity: Measurement and consequences for well-being, satisfaction and performance]. In R. Bouwen, K. De Witte, H. De Witte, & T. Taillieu (Eds.), *Van groep naar gemeenschap. Liber Amicorum Prof. Dr. Leo Lagrou* (pp. 325–350). Leuven: Garant.

Dias, J. G., & Vermunt, J. K. (2008). A bootstrap-based aggregate classifier for model-based clustering. *Computational Statistics*, *23*, 643–59.

Einarsen, S., Hoel, H., & Notelaers, G. (2009). Measuring exposure to bullying and harassment at work: Validity, factor structure and psychometric properties of the Negative Acts Questionnaire—Revised. *Work & Stress*, *23*, 24–44.

Gong, G., & Samaniego, F. J. (1981). Pseudo maximum likelihood estimation: Theory and applications. *The Annals of Statistics*, *76*, 861–889.

Goodman, L. A. (1974). The analysis of systems of qualitative variables when some of the variables are unobservable. Part I: A modified latent structure approach. *American Journal of Sociology*, *79*, 1179–1259.

Gourieroux, C., & Monfort, A. (1995). *Statistics and econometric models* (Vol. 2). Cambridge: Cambridge University Press.

Haberman, S. (1979). *Analysis of qualitative data. vol. 2: New developments*. New York: Academic Press.

Hagenaars, J. A. (1993). *Loglinear models with latent variables*. Newbury Park, CA: Sage.

Jöreskog, K. G., & Sörbom, D. (1986). *LISREL VI: Analysis of linear structural relationships by maximum likelihood and least squares methods*. Mooresville, IN: Scientific Software Inc.

Kam, J. A. (2011). Identifying changes in youth's subgroup membership over time based on their targeted communication about substance use with parents and friends. *Human Communication Research*, *37*, 324–349.

Lance, C. E., Cornwell, J. M., & Mulaik, S. A. (1988). Limited information parameter estimates for latent or mixed manifest and latent variable models. *Multivariate Behavioral Research*, *23*, 171–187.

Lanza, T. S., Tan, X., & Bray, C. B. (2013). Latent class analysis with distal outcomes: A flexible model-based approach. *Structural Equation Modeling*, *20*(1), 1–26.

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

Lu, I. R. R., & Thomas, D. R. (2008). Avoiding and correcting bias in score-based latent variable regression with discrete manifest items. *Structural Equation Modeling*, *15*, 462–490.

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

McCutcheon, A. L. (1985). A latent class analysis of tolerance for nonconformity in the American public. *Public Opinion Quarterly*, *494*, 474–488.

McCutcheon, A. L. (1987). *Latent class analysis*. Newbury Park, CA: Sage.

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide [Computer software manual]* (8th ed.). Los Angeles, CA: Muthen and Muthen.

Petersen, J., Bandeen-Roche, K., Budtz-Jørgensen, E., & Groes Larsen, K. (2012). Predicting latent class scores for subsequent analysis. *Psychometrika*, *77*, 244–262.

Ping, R. A. (1996). Latent variable interaction and quadratic effect estimation: A two-step technique using structural equation analysis. *Psychological Bulletin*, *119*, 166–175.

PSYCONES. (2006). *Psychological contracts across employment situations, final report*. DG Research, European Commission. Retrieved from http://cordis.europa.eu/documents/documentlibrary/100123961EN6.pdf

R Core Team. (2016). *R: A language and environment for statistical computing* [Computer software manual]. Vienna, Austria. https://www.R-project.org.

Skrondal, A., & Kuha, J. (2012). Improved regression calibration. *Psychometrika*, *77*, 649–669.

Skrondal, A., & Laake, P. (2001). Regression among factor scores. *Psychometrika*, *66*, 563–576.

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

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

Vermunt, J. K., & Magidson, J. (2016). *Technical guide for Latent GOLD 5.1: Basic, Advanced and Syntax*. Belmont, MA: Statistical Innovations.

Xue, Q. L., & Bandeen-Roche, K. (2002). Combining complete multivariate outcomes with incomplete covariate information: A latent class approach. *Biometrics*, *58*, 110–120.