Relating latent class assignments to external variables: standard errors for corrected inference

Z Bakk DL Oberski JK Vermunt

Dept of Methodology and Statistics, Tilburg University, The Netherlands

First author's address:

Zsuzsa Bakk

Department of Methodology and Statistics, FSW

Tilburg University

P.O. Box 90153

5000 LE Tilburg

The Netherlands

phone: +31 13 466 4030

email: z.bakk@uvt.nl

1 Introduction

Latent class analysis (LCA) is widely used in the social sciences to classify objects for further analysis (Ahlquist and Breunig, 2012). For example, McCutcheon (1985) examined the effect of education and age cohort on Americans' tolerance for nonconformity as obtained from a latent class model. Mustillo (2009, Table 4) provided a hard partitioning of new political parties in votatile party systems into "explosive," "contender," "flash," "flat," or "flop" but also noted (Table 5) that some classifications of parties as "flop" or otherwise are highly uncertain. Grimmer and Stewart (2013) discuss latent class analysis as an unsupervised machine learning method for political texts such as debates, legislation, news reports, and party manifestos; Grimmer (2013) related latent classes obtained from US Senators' press releases to their publicly expressed priorities.

While it is in principle preferable to model the relationship between indicators and classes and between classes and external variables simultaneously (Hagenaars, 1990, 1993; Bandeen-Roche, Miglioretti and Zegger, 1997), this practice has the disadvantages that classifications will be partially determined by the external variables, creating a circularity (Vermunt, 2010; Asparouhov and Muthén, 2012; Bakk, Tekle and Vermunt, in press), and that the estimation procedure can become cumbersomely complex. For these reasons, it is more common to follow a three-step procedure: in the first step, the latent class model is estimated, in the second step units are classified into classes using some assignment mechanism based on the first step, and in the third step the newly created observed variable is related to external variables using standard methods such as (logistic) regression. Bolck, Croon and Hagenaars (2004) noted that this procedure yields biased and inconsistent estimates of the latent class' relationship with external variables, and Vermunt (2010) provided an improved three-step procedure, in which the third step is amended by correcting for classification errors, removing the bias. Feingold, Tiberio and Capaldi (2013) applied this corrected method to the analysis of substance abuse and implementations in standard latent class software Mplus Version 7.1 (Asparouhov and Muthén, 2012) and Latent Gold Version 5 (Vermunt and Magidson, 2005) are available.

However, as we show in this article, the three-step procedure causes not only bias, but also additional variance in the estimates. This additional variance is not currently accounted for in standard software: even after the improved three-step procedures of Vermunt (2010) are applied, standard errors will be underestimated, and consequently confidence intervals will be too narrow and significance tests overly optimistic. Considering the broad applications of three-step latent class modelling, this is an undesirable situation. In this article we introduce two correction methods that can account for this additional uncertainty of the three-step estimates. We evaluate different possible estimators of the standard errors using Monte Carlo simulations, show how the optimal variance estimator depends on the class assignment method, and provide advice to correct inferences of applied research aiming to relate latent classes to external variables. Based on this study, the methods discussed have been made available to applied researchers in an experimental version of the standard latent class software Latent Gold."

Vermunt (2010) and Bakk, Tekle and Vermunt (in press) noted in simulation studies of the three-step latent class model that standard errors appeared to be underestimated. A similar phenomenon has been dealt with in the context of three-step structural equation modelling (Skrondal and Kuha, 2012; Oberski and Satorra, 2013), while econometric theory for two-stages least squares is already well-developed (Murphy and Topel, 1985). We apply the general theory (Gong and Samaniego, 1981) to latent class modelling, noting similarities and differences with these other approaches.

The structure of the paper is as follows: in Section 2 we introduce bias-adjusted three-step latent class analysis, while Section 3 introduces possible variance estimators of this model. Section 4 then evaluates and compares the performance of these different variance estimators in a simulation study. Section 5 illustrates their use in an application to tolerance for freedom of speech. We conclude in Section 6.

2 Bias-adjusted three-step latent class analysis

To model the relationship between a latent classification and external variables of interest without allowing the external variables to influence the classification, a three-step approach may be followed (Vermunt, 2010):

First step. Using only the indicator variables, estimate a latent class model;

Second step. Based on the first-step latent class model, create a new observed variable W that assigns to each unit its estimated latent class membership;

Third step. Relate the error-prone estimated classification W to external variables of interest while correcting for classification error in W.

We now discuss each of the steps and the resulting parameter uncertainty in turn.

2.1 First step: estimating a latent class model

The first step is a standard latent class analysis of K categorical indicator variables (McCutcheon, 1987; Goodman, 1974; Hagenaars, 1990). Given a sample of n units, the observations $\mathbf{Y}_i$ are modelled as arising from T unobserved (latent) classes X,

$$P(\mathbf{Y}_i) = \sum_{t=1}^{T} P(X = t) P(\mathbf{Y}_i | X = t). \tag{1}$$

The $T - 1$ unique latent class sizes (mixture proportions) will be denoted $P(X = t) = \rho_t$ and are the first set of parameters of the first-step model to be estimated.

The responses of each unit to the K categorical indicator variables are usually assumed to be locally independent given the unit's latent class membership. The conditional probability of the i-th response given the latent class can then be written as a product of conditional item responses,

$$P(\mathbf{Y}_i\,|X=t) = \prod_{k=1}^{K} P(Y_{ik}|X=t) = \prod_{k=1}^{K}\prod_{r=1}^{R_k} \pi_{ktr}^{I(Y_{ik}=r)}, \tag{2}$$

where the indicator variable $I(Y_{ik} = r) = 1$ if subject i has response r on item k, and 0 otherwise. The last step assumes that conditional item responses are equal for all units and defines the $(K-1)KT$ unique probabilities $\{\pi_{ktr}\}$ as the second set of first-step model parameters to be estimated.

The first-step log-likelihood of the sample data L_1 follows by combining equations 1 and 2 and assuming independence of observations:

$$L_1(\boldsymbol{\theta}_1) = \sum_{i=1}^{N} \log P(\mathbf{Y}_i) = \sum_{i=1}^{N} \log\left[\sum_{t=1}^{T} \rho_t \prod_{k=1}^{K}\prod_{r=1}^{R_k} \pi_{ktr}^{I(Y_{ik}=r)}\right]. \tag{3}$$

The first-step parameter vector to be estimated $\boldsymbol{\theta}_1 = [\boldsymbol{\rho}, \boldsymbol{\pi}]$ collects the latent class sizes $\boldsymbol{\rho}$ and conditional item response probabilities $\boldsymbol{\pi}$. Sample estimates $\hat{\boldsymbol{\theta}}_1$ of the first-step parameters can be obtained by maximum-likelihood (ML). Usually expectation-maximization, a quasi-Newton method, or a combination of both is used to maximize the first-step likelihood in Equation 3.

The maximum-likelihood estimates are sample estimates and will contain sampling variance. Assuming that the first-step model in Equation 3 is correct, standard theory suggests that the sampling variance equals the inverse of the Fisher information (negative of the Hessian matrix):

$$\boldsymbol{\Sigma}_1^H = (-\,\mathbf{H})^{-1}, \tag{4}$$

where the Hessian matrix $\mathbf{H}$ is defined as the second derivative of the first-step data log-likelihood with respect to the first-step parameters, $\mathbf{H} = \partial^2 L_1 / \partial \boldsymbol{\theta}_1 \, \partial \boldsymbol{\theta}_1'$.

The first-step model may not be correct–because the local independence assumption may not hold, for instance. If this misspecification is small, it will likewise have a small effect on the first-step estimates $\hat{\boldsymbol{\theta}}_1$. However, misspecification then still affects standard errors and sampling variance. The robust or "sandwich" variance should then be used,

$$\boldsymbol{\Sigma}_1^R = \boldsymbol{\Sigma}_1^H \, \mathbf{B} \, \boldsymbol{\Sigma}_1^H, \tag{5}$$

where the "meat" of the sandwich, $\mathbf{B}$, is the average outer product of the casewise gradients (White, 1982). Although the robust variance estimator corrects for model misspecification, it will also lead to a loss of efficiency (Kauermann and Carroll, 2001). It is therefore not clear in practice whether $\boldsymbol{\Sigma}_1^H$ or $\boldsymbol{\Sigma}_1^R$ should be preferred.

2.2 Second step: assignment of units to classes

After estimation of the latent class model in the first step, a new variable W is created, assigning each unit to an estimated class. Following Bayes' rule, each unit's posterior probability of belonging to class t is

$$P(X = t|\mathbf{Y}_i) = \frac{P(X = t)P(\mathbf{Y}_i|X = t)}{P(\mathbf{Y}_i)}. \tag{6}$$

Sample estimates of the posterior probabilities $P(X = t|\mathbf{Y}_i)$ can be obtained by replacing class sizes $P(X = t)$ with $\hat{\rho}_t$, conditional probability $P(\mathbf{Y}_i|X = t)$ with $\prod \hat{\pi}_{ktr}$, and generally substituting elements of $\boldsymbol{\theta}_1$ in Equation 6 with their first-step sample estimates $\hat{\boldsymbol{\theta}}_1$.

The sample estimates of the posterior class probabilities from Equation 6 can be used in different ways to create an estimated class membership variable W (Vermunt, 2010). We discuss the two most widely known and applied assignment rules: modal and proportional assignment.

The modal assignment rule to generate a posterior classification W is the most widely used (Collins and Lanza, 2010, p. 72). Each unit is simply assigned the class label with the largest (modal) estimated posterior probability from Equation 6. Using modal assignment the value of $P(W_i = t|\mathbf{Y}_i) = 1$ is assigned for $P(X = t|\mathbf{Y_i}) > P(X = t^{'}|\mathbf{Y_i})$ for all $t \neq t^{'}$. For all other classes this value is set to 0, leading to a hard partitioning.

Proportional assignment, in contrast, is a soft partitioning method (Dias and Vermunt, 2008). For each unit, T records are first created, one for each latent class. The T values of W_i are then set equal to the posterior probabilities $P(X = t|\mathbf{Y}_i)$. The data matrix is therefore expanded to include T instead of one records for each of the n units, where the within-unit values of the class assignment variable W will act as weights in the third step of the analysis.

Irrespective of the assignment method used, the true (X) and assigned (W) class membership scores will differ. Classification errors are therefore always present, even if the entire population were observed. The amount of classification errors will depend on the posterior classification and the assignment method chosen. After assignment, the assignment variable W will require correction for classification errors in the third step; therefore, the amount of error in it must first be calculated (Bolck, Croon and Hagenaars, 2004).

Summing over all observed data patterns the amount of classification errors can be expressed as the posterior class membership conditional on the true value (Vermunt, 2010;

Bakk, Tekle and Vermunt, in press),

$$P(W = s|X = t) \;=\; \frac{\frac{1}{N}\sum_{i=1}^{N} P(X = t|\mathbf{Y_i})P(W_i = s|\mathbf{Y}_i)}{P(X = t)}. \tag{7}$$

Note that while for any assignment method used the general form of equation 7 is the same, the values of $P(W_i = s|\mathbf{Y}_i)$ will differ per assignment method, and thus the amount of classification error $P(W = s|X = t)$ will also differ by assignment method. For example, $P(W_i = s|\mathbf{Y}_i)$ is either 0 or 1 using modal assignment, and with proportional assignment $P(W_i = s|\mathbf{Y}) = P(X = s|\mathbf{Y})$.As we will show later this difference is not problematic, it just reflects that the amount of classification error depends on the assignment method used.

The classification error can be re-expressed on the logit scale as follows:

$$P(W = s|X = t) = \frac{\exp(\gamma_{st})}{\sum_{s=1}^{T}\exp(\gamma_{st})}, \tag{8}$$

where

$$\gamma_{st} = \log\left[\frac{P(W = s|X = t)}{P(W = t|X = t)}\right].$$

Note that the logistic γ_{st} parameters do not constitute free parameters but follow as a function of the first-step results and the assignment rule chosen.

We collect the γ_{st} parameters in the vector $\boldsymbol{\theta}_2$, with sample estimates $\hat{\boldsymbol{\theta}}_2$, calculated directly from $\hat{\boldsymbol{\theta}}_1$. These logistic effects of the true latent class on the estimated classification W are later needed to correct for classification error. Since these logit coefficients are calculated from the uncertain first-step estimates, they are themselves uncertain. Their sampling variance $\boldsymbol{\Sigma}_2$ can be obtained using the delta method from the variance of the first step model: (Oehlert, 1992)

$$\Sigma_2 = \left(\frac{\partial\theta_2}{\partial\theta_1}\right)\Sigma_1\left(\frac{\partial\theta_2}{\partial\theta_1}\right)' \tag{9}$$

Either of the Σ_1 estimators discussed above can be plugged in to the formula, leading to an observed Hessian based (Σ_2^H) or robust (Σ_2^R) variance estimator of the second step parameters.

2.3 Third step: relating estimated membership to covariates

In the third step the assigned classification W is related to a vector of covariates, $\mathbf{Z}$, say, while also correcting for classification error in W. Logistic regression of W on $\mathbf{Z}$ may appear to be an obvious solution, but would yield biased estimates due to classification errors in W. In effect, the relationship with the error-prone W is modelled, where the relationship with the true but unobserved latent class variable X is of interest, leading to measurement error effects on the parameter estimates (Bolck, Croon and Hagenaars, 2004).

Bolck, Croon and Hagenaars (2004) showed how the $P(X = t|\mathbf{Z}_i)$, and $P(W = s|\mathbf{Z}_i)$ are related to each other, namely that the $P(W = s|\mathbf{Z}_i)$ can be written as a weighted sum of the latent classes given the covariates,with the classification error probabilities as the weights:

$$P(W = s|\mathbf{Z}_i) = \sum_{t=1}^{T} P(X = t|\mathbf{Z}_i)P(W = s|X = t). \tag{10}$$

Vermunt (2010) noted that Equation 10 can be seen as a latent class model with W as a single indicator and fixed “known” classification error probabilities $P(W = s|X = t)$. This means that relating the estimated membership to covariates while correcting for classification errors can be achieved by using standard latent class software that allows the user to fix classification error parameters to those obtained in the second step.

This model is composed of two parts: (1) the structural part, i.e. the model of interest for $P(X = t|\mathbf{Z}_i)$, relating the latent class membership to the vector of external variables and (2) the measurement part $P(W = s|X = t)$ fixed to the parameter values estimated in step 2, as shown in Equation 7.

Denoting by Z_{iq} the value of subject i on one of the Q covariates, the structural part of the model can be parametrized by means of a multinomial logistic regression model,

$$P(X = t|\mathbf{Z}_i) = \frac{\exp(\beta_{0t} + \sum_{q=1}^{Q} \beta_{qt} Z_{iq})}{\sum_{s=1}^{T} \exp(\beta_{0t} + \sum_{q=1}^{Q} \beta_{qs} Z_{iq})}. \tag{11}$$

Although we only present the third-step model with predictors of latent class membership here, Bakk, Tekle and Vermunt (in press) showed how the correction method can be used for a wider class of models, including models where the class membership is a predictor of a distal outcome variable, or with multiple latent variables. For the measurement part the logistic parametrization can be used as defined in equation 8.

The parameters of interest are the logistic regression coefficients β_{qt}, gathered in the vector $\boldsymbol{\theta}_3$. Consistent estimates $\hat{\boldsymbol{\theta}}_3$ can be obtained by maximizing the third-step log-likelihood (Vermunt, 2010),

$$L_3(\boldsymbol{\theta}_3 \,|\, \boldsymbol{\theta}_2 = \hat{\boldsymbol{\theta}}_2) = \sum_{n=1}^{N} \sum_{s=1}^{T} P(W = s|\mathbf{Y}_i) \log \sum_{t=1}^{T} P(X = t|\mathbf{Z}_i) P(W = s|X = t). \tag{12}$$

Thus, in the third step, the logistic regression coefficients, contained in the third-step parameter vector $\boldsymbol{\theta}_3$, are freely estimated, while the classification errors of the class membership variable W as a measure of X, contained in the second-step parameter vector $\boldsymbol{\theta}_2$, are held fixed at their sample maximum-likelihood estimates, $\boldsymbol{\theta}_2 = \hat{\boldsymbol{\theta}}_2$. The third-step

ML estimates can therefore be seen as conditional estimates $(\hat{\boldsymbol{\theta}}_3 | \, \boldsymbol{\theta}_2 = \hat{\boldsymbol{\theta}}_2)$.

3 Variance of the third-step estimates

Although the third-step maximum-likelihood estimates $\hat{\boldsymbol{\theta}}_3$ are consistent, their sampling variance now contains two sources of variation: that variation due to estimation at the third step, and that carried over from the first step. Ignoring the second source of variance will lead to an underestimation of the standard errors, as the results of previous simulation studies showed (Vermunt, 2010; Bakk, Tekle and Vermunt, in press).

To see why underestimation occurs, write the variance of the third-step estimate as conditional on the second step (Oberski and Satorra, 2013):

$$\boldsymbol{\Sigma}_3^* \equiv \mathrm{Var}(\hat{\boldsymbol{\theta}}_3) = \mathrm{E}_{\theta_2}[\mathrm{Var}(\hat{\boldsymbol{\theta}}_3 | \, \boldsymbol{\theta}_2)] + \mathrm{Var}_{\theta_2}[\mathrm{E}(\hat{\boldsymbol{\theta}}_3 | \, \boldsymbol{\theta}_2)]. \tag{13}$$

The first term in Equation 13 corresponds approximately to the usual variance calculations obtained after fixing parameters in the third step,

$$\mathrm{E}_{\theta_2}[\mathrm{Var}(\hat{\boldsymbol{\theta}}_3 | \, \boldsymbol{\theta}_2)] \approx \boldsymbol{\Sigma}_3, \tag{14}$$

where $\boldsymbol{\Sigma}_3$ may, again, be estimated as the inverse third-step Fisher information or with the robust variance estimator. This is the basis for standard errors currently given by standard latent class analysis software when performing three-step analysis.

In the case of proportional assignment, each unit has several cases associated with it. Simulation studies by Vermunt (2010) and Bakk, Tekle and Vermunt (in press) found that using the third-step Hessian matrix to obtain an estimator of $\boldsymbol{\Sigma}_3$, standard errors were underestimated for modal assignment but overestimated for proportional assignment, a phenomenon that can be explained by the duplication of records present in proportional assignment. To correct the standard errors for this duplication, $\mathbf{\Sigma}_3$ must be estimated with the well-known "complex sampling" (clustered) robust variance estimator (Michel, Hofstede and Steenkamp, 1998), which will be denoted $\mathbf{\Sigma}_3^R$. Using these estimator we expect the standard error estimates to be down-weighted, because the sum of square of weights is always smaller then one with proportional assignment.

The second term in Equation 13 can be obtained by a first-order Taylor expansion (Gong and Samaniego, 1981; Oberski and Satorra, 2013),

$$\mathrm{Var}_{\theta_2}[\mathrm{E}(\hat{\boldsymbol{\theta}}_3 | \, \boldsymbol{\theta}_2)] \approx \left(\frac{\partial \, \boldsymbol{\theta}_3}{\partial \, \boldsymbol{\theta}_2}\right) \mathbf{\Sigma}_2 \left(\frac{\partial \, \boldsymbol{\theta}_3}{\partial \, \boldsymbol{\theta}_2}\right)', \tag{15}$$

where an estimate of $\mathbf{\Sigma}_2$ is available from the second step, and $\partial \, \boldsymbol{\theta}_3 \, / \partial \, \boldsymbol{\theta}_2$ can be obtained using implicit function theorem:

$$\frac{\partial \, \boldsymbol{\theta}_3}{\partial \, \boldsymbol{\theta}_2} = \left(-\frac{\partial^2 L_3}{\partial \, \boldsymbol{\theta}_3 \, \partial \, \boldsymbol{\theta}_3'}\right)^{-1} \frac{\partial^2 L_3}{\partial \, \boldsymbol{\theta}_3 \, \partial \, \boldsymbol{\theta}_2'} \equiv - \, \mathbf{H}_3^{-1} \, \mathbf{C}, \tag{16}$$

which thus requires obtaining the second derivatives of the third-step log likelihood towards the free parameters ($\mathbf{H}$) and towards the free parameters with respect to the fixed parameters ($\mathbf{C}$). Therefore, the third-step variance defined in equation 13 can be written as the sum of two positive-definite terms,

$$\mathbf{\Sigma}_3^* = \mathbf{\Sigma}_3 + \mathbf{H}_3^{-1} \, \mathbf{C} \, \mathbf{\Sigma}_2 \, \mathbf{C}' \, \mathbf{H}_3^{-1} \, . \tag{17}$$

If a second-order Taylor expansion is used instead of Equation 15, an additional term results (Gong and Samaniego, 1981, Theorem 2.2), leading to

$$\mathbf{\Sigma}_3^{**} = \mathbf{\Sigma}_3 + \mathbf{H}_3^{-1} (\mathbf{C} \, \mathbf{\Sigma}_2 \, \mathbf{C}' - \mathbf{C} \, \mathbf{H}_2^{-1} \, \mathbf{R}' - \mathbf{R} \, \mathbf{H}_2^{-1} \, \mathbf{C}') \, \mathbf{H}_3^{-1}, \tag{18}$$

where the $\mathbf{R}$ matrix is the outer product of the case-wise gradients of the first and third-step models, $\mathbf{R} = (\partial L_3/\partial \boldsymbol{\theta}_3)'(\partial L_1/\partial \boldsymbol{\theta}_1)$; its precise form for the latent class model is given in the appendix. However, perhaps surprisingly, this extra term vanishes as the sample size increases: provided the first-step estimates are consistent, asymptotically $\mathbf{R} = \mathbf{0}$ (Parke, 1986). Therefore, the two variance estimators are equal in large samples, $\boldsymbol{\Sigma}^{**} \stackrel{a}{=} \boldsymbol{\Sigma}^{*}$, although they may not be equal in small samples. In small samples it is possible that $\boldsymbol{\Sigma}^{*}$ will overestimate the standard errors of the third-step estimates, although this overestimation should decrease with sample size; on the other hand, the calculation of the extra terms in $\boldsymbol{\Sigma}^{**}$ may add considerable effort and instability to the standard errors.

Whether $\boldsymbol{\Sigma}^{*}$ or $\boldsymbol{\Sigma}^{**}$ is the more appropriate variance estimate is therefore unclear. Furthermore, it can be concluded from the preceding discussion that at each step a range of possible choices of variance estimators exists. The following section investigates how combinations of these different choices perform and which, if any, of the standard error corrections is likely to be necessary in practice.

4 Monte Carlo simulation

4.1 Design

In order to see which variance estimator performs the best, we crossed the choice of variance estimators (for $\boldsymbol{\Sigma}_2$ and $\boldsymbol{\Sigma}_3$: observed Hessian based or robust) with the options for correcting for uncertainty ($\boldsymbol{\Sigma}_3$ - uncorrected, $\boldsymbol{\Sigma}_3^{*}$ first order and $\boldsymbol{\Sigma}_3^{**}$ second order correction) for both modal and proportional assignment. In the following table we summarize the different choices of the variance estimators compared.

As used in Table 1 the 1^{st} order correction, $\boldsymbol{\Sigma}_3^{*}$ is defined in equation 17 and the 2^{nd} order correction, $\boldsymbol{\Sigma}_3^{**}$ in 18, and $\boldsymbol{\Sigma}_3$ is the variance of the free parameters ignoring the additional uncertainty attributable to the fixed parameter values. In reporting the simulation

Table 1: Possible variance estimators of the third step model

Final	Components	
	2^{nd} step	3^{rd} step
Uncorrected ($\mathbf{\Sigma}_3$)	-	Hessian ($\mathbf{\Sigma}_3^H$)
	-	Robust ($\mathbf{\Sigma}_3^R$)
1^{st} order correction ($\mathbf{\Sigma}_3^*$)	Hessian ($\mathbf{\Sigma}_2^H$)	Hessian ($\mathbf{\Sigma}_3^H$)
	Hessian($\mathbf{\Sigma}_2^H$)	Robust ($\mathbf{\Sigma}_3^R$)
	Robust($\mathbf{\Sigma}_2^R$)	Robust ($\mathbf{\Sigma}_3^R$)
2^{nd} order correction ($\mathbf{\Sigma}_3^{**}$)	Hessian($\mathbf{\Sigma}_2^H$)	Hessian ($\mathbf{\Sigma}_3^H$)
	Hessian ($\mathbf{\Sigma}_2^H$)	Robust ($\mathbf{\Sigma}_3^R$)
	Robust ($\mathbf{\Sigma}_2^R$)	Robust ($\mathbf{\Sigma}_3^R$)

study results and real data example we use the term $\mathbf{\Sigma}_3^R$ in case of proportional assignment for the complex sampling variance estimator (Michel, Hofstede and Steenkamp, 1998), and for modal assignment for the sandwich estimator as defined by White (1982). All in all we investigate 8 variance estimators for each of the two assignment methods separately.

The need for the uncertainty correction is expected to depend on the amount of uncertainty about the model parameters, that we varied by changing sample size and separation between classes.

As population model we chose a LCA model with 3 classes measured by 6 dichotomous indicators, and regressed on 3 numerical covariates (each with five categories: 1-5). The first class is likely to give positive response on all 6 items, class two has a high probability of a positive response on the first 3 items, and negative response on the other three items. In class three all items have a high probability of a negative answer. We manipulated the separation between classes by changing the size of the conditional probability of the indicators given the classes. The two levels of separation we used for the probability of a positive answer are .80 and .90, corresponding to entropy R^2 values of .65 and .90. We chose the following sample sizes: 500, 1000, 2000. Thus in total we had 6 conditions of combinations of sample size and separation between classes, in which the performance of all 8 variance estimators was compared for both modal and proportional assignment.For each condition 500 replications were used.

Using the first class as reference category we set the logit parameters of covariate effects on latent classes to -2(β_{12})and 1 (β_{13}) for the effect of Z_1 on X, to 1 (β_{22}) and 0 (β_{23}) for the effect of Z_2 on X, and to 0 for both parameters (β_{32}, β_{33}) for the effect of Z_3 on X. The intercepts were set to values yielding equal class sizes.

Two measures were used to compare the performance of the variance estimators. We compared the coverage rate over replications to a nominal 95 percent rate, and the average standard errors (se) across replications to the standard deviation (sd) across replications. For a well performing standard error estimator we expect the se/sd to be 1. Also the coverage rate should be 95 %, which is the nominal coverage rate used.

We used the computer programs Latent GOLD (Vermunt and Magidson, 2005) and R (Venables, Smith and the R Core Team, 2013) to run the analysis.

4.2 Simulation Results

First we compare the parameter estimates and standard deviation across replications obtained with the three-step approach with the two assignment methods and the one-step approach in order to see whether the three-step estimates are comparable with regard to parameter bias and efficiency to the one-step approach. In Table 2 we report the mean parameter estimates over all replications, and the standard deviation across replications for all three estimation methods. On average the parameter bias is low with all three estimators for all the parameters. We compared the efficiency of the parameter estimators by comparing the standard deviation across replications. As we can see in Table 2 the standard deviations of all parameters are very close to each other with the three methods.These results are in accordance with previous simulation studies (Bakk, Tekle and Vermunt, in press; Vermunt, 2010), and show that the three- step approach can be used

Table 2: Parameter estimates and their standard deviation (sd) for all parameters averaged over all conditions for all estimators

		Modal		Proportional		One-Step	
Value	True	Estimate	sd	Estimate	sd	Estimate	sd
β_{12}	-2.00	-1.98	0.30	-1.97	0.28	-2.07	0.30
β_{13}	1.00	1.00	0.12	1.00	0.11	1.01	0.11
β_{22}	1.00	1.00	0.17	0.98	0.16	1.02	0.18
β_{23}	0.00	0.00	0.08	0.00	0.07	0.00	0.08
β_{32}	0.00	0.00	0.11	0.00	0.11	0.00	0.11
β_{33}	0.00	0.00	0.07	0.00	0.07	0.00	0.07

without loss of efficiency or parameter bias.

Given the unbiased parameter estimates reported in Table 2 in the following we restrict the discussion only to the variance estimators of the third-step model.

Let us first look on the results averaged across all conditions of sample size and separation between classes, that are reported in Table 3 for one parameter ($\beta_{13} = 1.00$). The results for the other parameters are very similar.

For modal assignment, as can be seen in Table 3 the two uncorrected standard error estimators that do not account for the additional uncertainty (Σ_3^H and Σ_3^R) underestimate the variance (the se/sd is .95 for Σ_3^H, and .97 for Σ_3^R). Using either of the correction methods ($\boldsymbol{\Sigma}_3^*$ or $\boldsymbol{\Sigma}_3^{**}$) improves the results for both Hessian based and robust estimator. Comparing the first and second order corrections ($\boldsymbol{\Sigma}_3^*$, $\boldsymbol{\Sigma}_3^{**}$) to each other we see that the standard error estimates obtained with the later are slightly higher irrespective of the choice for robust or Hessian based estimator. When comparing observed Hessian based to robust estimator for the modal assignment, we see that the standard errors obtained with the later are somewhat larger, thus less efficient. The differences are although small, as can be seen from the coverage rate, which is almost the same with all estimators.

Next, looking on the results for proportional assignment, we see, that as hypothesized the standard error estimates obtained with the observed Hessian based estimator overestimate the standard error for all three estimators ($\boldsymbol{\Sigma}_3$, $\boldsymbol{\Sigma}_3^*$, $\boldsymbol{\Sigma}_3^{**}$) This can be seen from both

Table 3: Comparison of the different variance estimators averaged across all conditions for one parameter, β_{13} for modal and proportional assignment separately

Final	Components		Modal			Proportional		
	2^{nd} step	3^{rd} step	se	se/sd	coverage	se	se/sd	coverage
$\mathbf{\Sigma}_3$	-	$\mathbf{\Sigma}_3^H$	0.11	0.95	0.95	0.12	1.08	0.97
$\mathbf{\Sigma}_3^*$	$\mathbf{\Sigma}_2^H$	$\mathbf{\Sigma}_3^H$	0.12	1.03	0.96	0.13	1.14	0.98
$\mathbf{\Sigma}_3^{**}$	$\mathbf{\Sigma}_2^H$	$\mathbf{\Sigma}_3^H$	0.12	1.04	0.96	0.13	1.12	0.97
$\mathbf{\Sigma}_3$	-	$\mathbf{\Sigma}_3^R$	0.12	0.97	0.95	0.11	0.99	0.95
$\mathbf{\Sigma}_3^*$	$\mathbf{\Sigma}_2^H$	$\mathbf{\Sigma}_3^R$	0.12	1.04	0.96	0.12	1.05	0.96
$\mathbf{\Sigma}_3^{**}$	$\mathbf{\Sigma}_2^H$	$\mathbf{\Sigma}_3^R$	0.13	1.05	0.96	0.12	1.03	0.96
$\mathbf{\Sigma}_3^*$	$\mathbf{\Sigma}_2^R$	$\mathbf{\Sigma}_3^R$	0.13	1.05	0.96	0.12	1.06	0.96
$\mathbf{\Sigma}_3^{**}$	$\mathbf{\Sigma}_2^R$	$\mathbf{\Sigma}_3^R$	0.13	1.06	0.96	0.12	1.04	0.96

Note: $\mathbf{\Sigma}_3^*$ is the 1^{st} and $\mathbf{\Sigma}_3^{**}$ the 2^{nd} order correction, as defined in equation 17 and 18, and $\mathbf{\Sigma}^H$ and $\mathbf{\Sigma}^R$ are the Hessian based and robust estimators

the se/sd (which is higher then 1 for all three estimators) and coverage rate (that is .97 for $\mathbf{\Sigma}_3$ and $\mathbf{\Sigma}_3^{**}$ and .98 for $\mathbf{\Sigma}_3^*$). Using the robust variance estimator in the third step improves the results (the se/sd using $\mathbf{\Sigma}_3^R$ is .99, and using the correction methods this value becomes slightly larger then 1). Similarly to modal assignment we can see, that using robust variance estimator in the first step yields larger standard error estimates.

Following we look separately into the results averaged over the different levels of separation between classes and the different sample size conditions.

First the results averaged over the three sample sizes separately for the 2 separation levels are presented. For the condition with high separation between the classes (entropy R^2=.90) all variance estimators perform well. In case of modal assignment all the variance estimators obtaining se/sd between 1.00-1.02, and coverage rate of 96 %. The same holds for all standard error estimates for proportional assignment that are based on the robust variance estimator. As such we do not present these results in more detail, but move toward the discussion of the low separation condition, where we see more variability.

In Table 4 the results averaged over the three sample sizes for the low separation condition are presented. For modal assignment the uncertainty uncorrected standard er-

Table 4: Comparison of the different variance estimators across the three sample sizes,for the low separation levels for one parameter, β_{13} for modal and proportional assignment separately

Final	Components		Modal			Proportional		
	2^{nd} step	3^{rd} step	se	se/sd	coverage	se	se/sd	coverage
$\mathbf{\Sigma}_3$	-	$\mathbf{\Sigma}_3^H$	0.12	0.91	0.93	0.14	1.11	0.98
$\mathbf{\Sigma}_3^*$	$\mathbf{\Sigma}_2^H$	$\mathbf{\Sigma}_3^H$	0.14	1.03	0.96	0.15	1.21	0.98
$\mathbf{\Sigma}_3^{**}$	$\mathbf{\Sigma}_2^H$	$\mathbf{\Sigma}_3^H$	0.14	1.05	0.96	0.15	1.18	0.98
$\mathbf{\Sigma}_3$	-	$\mathbf{\Sigma}_3^R$	0.13	0.93	0.94	0.12	0.97	0.95
$\mathbf{\Sigma}_3^*$	$\mathbf{\Sigma}_2^H$	$\mathbf{\Sigma}_3^R$	0.14	1.05	0.96	0.14	1.08	0.97
$\mathbf{\Sigma}_3^{**}$	$\mathbf{\Sigma}_2^H$	$\mathbf{\Sigma}_3^R$	0.14	1.06	0.96	0.13	1.05	0.96
$\mathbf{\Sigma}_3^*$	$\mathbf{\Sigma}_2^R$	$\mathbf{\Sigma}_3^R$	0.15	1.07	0.96	0.14	1.10	0.97
$\mathbf{\Sigma}_3^{**}$	$\mathbf{\Sigma}_2^R$	$\mathbf{\Sigma}_3^R$	0.15	1.08	0.96	0.14	1.07	0.96

Note: $\mathbf{\Sigma}_3^*$ is the 1^{st} and $\mathbf{\Sigma}_3^{**}$ the 2^{nd} order correction, as defined in equation 17 and 18, and $\mathbf{\Sigma}^H$ and $\mathbf{\Sigma}^R$ are the Hessian based and robust estimators

ror estimate ($\mathbf{\Sigma}_3$) underestimates the standard error with both Hessian based and robust estimators (se/sd is .91 and .93 while coverage rate is .93 and .94 for Σ_3^H and Σ_3^R respectively). Using either of the correction methods ($\mathbf{\Sigma}_3^*, \mathbf{\Sigma}_3^{**}$) the se/sd becomes slightly larger then 1, and the coverage rate increases to .96 %, for both the Hessian based and robust estimators. When comparing the observed Hessian based estimates to the robust estimates we can see that the later obtains slightly larger standard error estimates.

For proportional assignment we can see that the variance estimators that use the observed Hessian in the third-step overestimate the standard error. Using the robust variance estimator in the third step decreases the variance. Once this is used, the difference between the standard error estimates is small ($\mathbf{\Sigma}_3$ obtains se/sd 0.97,with $\mathbf{\Sigma}_3^*$ this is 1.08, and with $\mathbf{\Sigma}_3^{**}$ 1.05). Using robust variance estimator in the first step increases the standard error estimates.

Next in Table 5 we present the standard error estimates averaged over the two separation levels separately for the three sample size conditions. For modal assignment we can see that in the small sample size condition the uncorrected standard error estimates are underestimated (se/sd 0.87 for Σ_3^H and 0.90 for Σ_3^R and coverage rate 0.92 and 0.93 respectively), but using any of the correction methods this values get closer to 1. Comparing the first and second order correction ($\mathbf{\Sigma}_3^*$, $\mathbf{\Sigma}_3^{**}$) we see that the standard error estimates obtained with the later are slightly larger irrespective of the choice of $\mathbf{\Sigma}_3$. The same tendencies can be seen in the larger sample size conditions as well, though in the 2000 sample size condition it can be seen, that on average the uncorrected standard error estimates are the same as the corrected ones with the precision of 2 decimals. Comparing the Hessian based estimators to the robust ones we see that the later ones are somewhat larger.

Using proportional assignment the same tendencies can be observed, once in the third step the robust standard error is used.

In summary it can be said, that in conditions where the uncertainty about the fixed parameters is high (that is low separation between classes and/or low sample sizes) the use of the uncertainty correction is needed. It can be seen that the difference with the results using $\mathbf{\Sigma}_3^*$ and $\mathbf{\Sigma}_3^{**}$ is low, thus the use of the first order correction is recommended, because it needs less calculations. With regard to the choice of Hessian or robust variance estimator we see that in case of proportional assignment this choice is important. With proportional assignment the use of robust estimator is recommended for all situations, while for modal assignment this choice is not so relevant.

5 Application to tolerance for freedom of speech

We illustrate the use of the different variance estimators for the three step LCA using data from the 2004 General Social Survey. We built an LC model with 5 indicator variables measuring the respondents attitude toward freedom of speech of different groups. The 5 groups considered are: anti-religionists, communists, homosexuals, militarists, racists.

Table 5: Comparison of the different variance estimators averaged across the two separation levels,for the 3 sample sizes for one parameter, β_{13} for modal and proportional assignment separately

Final	Components		500			1000			2000		
	2^{nd} step	3^{rd} step	se	se/sd	coverage	se	se/sd	coverage	se	se/sd	coverage
						Modal					
$\mathbf{\Sigma}_3$		$\mathbf{\Sigma}_3^H$	0.16	0.87	0.92	0.11	1.01	0.95	0.08	1.05	0.96
$\mathbf{\Sigma}_3^*$	$\mathbf{\Sigma}_2^H$	$\mathbf{\Sigma}_3^H$	0.17	0.96	0.95	0.12	1.08	0.97	0.08	1.13	0.97
$\mathbf{\Sigma}_3^{**}$	$\mathbf{\Sigma}_2^H$	$\mathbf{\Sigma}_3^H$	0.17	0.97	0.95	0.12	1.09	0.97	0.08	1.12	0.97
$\mathbf{\Sigma}_3$		$\mathbf{\Sigma}_3^R$	0.16	0.90	0.93	0.11	1.02	0.95	0.08	1.06	0.96
$\mathbf{\Sigma}_3^*$	$\mathbf{\Sigma}_2^H$	$\mathbf{\Sigma}_3^R$	0.18	0.98	0.95	0.12	1.08	0.97	0.08	1.11	0.97
$\mathbf{\Sigma}_3^{**}$	$\mathbf{\Sigma}_2^H$	$\mathbf{\Sigma}_3^R$	0.18	0.99	0.95	0.12	1.08	0.97	0.08	1.13	0.97
$\mathbf{\Sigma}_3^*$	$\mathbf{\Sigma}_2^R$	$\mathbf{\Sigma}_3^R$	0.18	1.01	0.95	0.12	1.10	0.97	0.08	1.11	0.97
$\mathbf{\Sigma}_3^{**}$	$\mathbf{\Sigma}_2^R$	$\mathbf{\Sigma}_3^R$	0.18	1.01	0.95	0.12	1.09	0.97	0.08	1.13	0.97
						Proportional					
$\mathbf{\Sigma}_3$		$\mathbf{\Sigma}_3^H$	0.17	0.99	0.96	0.12	1.15	0.98	0.08	1.2	0.98
$\mathbf{\Sigma}_3^*$	$\mathbf{\Sigma}_2^H$	$\mathbf{\Sigma}_3^H$	0.18	1.06	0.96	0.12	1.20	0.98	0.09	1.25	0.98
$\mathbf{\Sigma}_3^{**}$	$\mathbf{\Sigma}_2^H$	$\mathbf{\Sigma}_3^H$	0.18	1.04	0.96	0.12	1.18	0.98	0.09	1.23	0.98
$\mathbf{\Sigma}_3$		$\mathbf{\Sigma}_3^R$	0.16	0.92	0.94	0.11	1.04	0.96	0.08	1.08	0.96
$\mathbf{\Sigma}_3^*$	$\mathbf{\Sigma}_2^H$	$\mathbf{\Sigma}_3^R$	0.17	0.99	0.95	0.11	1.10	0.97	0.08	1.14	0.97
$\mathbf{\Sigma}_3^{**}$	$\mathbf{\Sigma}_2^H$	$\mathbf{\Sigma}_3^R$	0.17	0.97	0.95	0.11	1.08	0.96	0.08	1.12	0.97
$\mathbf{\Sigma}_3^*$	$\mathbf{\Sigma}_2^R$	$\mathbf{\Sigma}_3^R$	0.18	0.98	0.95	0.11	1.09	0.97	0.08	1.14	0.97
$\mathbf{\Sigma}_3^{**}$	$\mathbf{\Sigma}_2^R$	$\mathbf{\Sigma}_3^R$	0.18	1.00	0.95	0.11	1.10	0.97	0.08	1.12	0.97

Note: $\mathbf{\Sigma}_3^*$ is the 1^{st} and $\mathbf{\Sigma}_3^{**}$ the 2^{nd} order correction, as defined in equation 17 and 18, and $\mathbf{\Sigma}^H$ and $\mathbf{\Sigma}^R$ are the Hessian based and robust estimators

The question related to anti-religion is formulated as follows: 'There are always some people whose ideas are considered bad or dangerous by other people. For instance somebody who is against all churches and religion ... If such a person wanted to make a speech in your city/town/communnity against churches and religion, should he be allowed to speak or not'. The two possible answers are 'allowed' and 'not allowed'. The other items have the same formulation. As explanatory variables we use (years of) education, race (1=white, 2=black, 3=other) and gender (1=male, 2=female). We choose to exclude all cases with missing data, in this way we obtained a sample size of 852.

A three-class model fitted the data the best. This model was selected by both AIC and BIC, the residuals in all two-way tables were small, and the likelihood and Pearson

Table 6: Parameter estimates of the 3 class model for tolerance for freedom of speech: class sizes and class specific probabilities of agreeing with the indicator concerned

	Intolerant	Neutral	Tolerant
Class sizes	0.15	0.32	0.53
Racist	0.12	0.46	0.90
Homosexuals	0.32	0.85	0.99
Communists	0.05	0.59	0.98
Atheists	0.16	0.75	0.99
Militarists	0.03	0.54	0.96

chi-squared statistics were borderline significant. The three classes that were found were nicely ordered: the smallest group (15 %) is against freedom of speech for all groups, the second largest (32%) is neutral and the largest class (53%) is in favour of freedom of speech. In Table 6 the LC solution is presented.

Next the latent class variable was regressed on education, race and gender using the one-step method and the bias-adjusted three-step method with modal and proportional assignment, using the different standard error options discussed. We excluded the standard error options that use robust estimator in the first step, because the simulation study showed that these estimator is the least efficient. In Table 7 we report the estimated effect of covariates on the class membership and their standard errors obtained with the different estimators.

Class 1 (intolerant) serves as the baseline, and white and male are used as reference categories for race and gender. The parameter estimates show that the three-step approaches yield estimates which are close to the ones obtained with one-step ML estimation. Overall it seems that the parameter estimates obtained with proportional assignment are closer to one-step ML than the estimates obtained with modal assignment. The results show that more educated people tend to be more tolerant. Black people and people of other colour tend to have a lower probability to belong to the tolerant group then white people. Females have a smaller probability to belong to the neutral and tolerant groups

Table 7: The effect of covariates on LC's using dummy coding, with class 1 as reference category: Parameter and standard error estimates

	EducCl2	EducCl3	SexCl2	SexCl3	BlackCl2	BlackCl3	OtherCl2	OtherCl3
Modal assignment								
Estimate	0.07	0.40	-0.31	-0.35	0.02	-0.77	0.33	-0.90
se($\mathbf{\Sigma}_3^{H}$)	0.05	0.05	0.29	0.25	0.37	0.35	0.50	0.52
se($\mathbf{\Sigma}_3^{*H}$)	0.06	0.05	0.29	0.25	0.37	0.36	0.51	0.53
se($\mathbf{\Sigma}_3^{**H}$)	0.06	0.06	0.29	0.25	0.37	0.36	0.52	0.53
se($\mathbf{\Sigma}_3^{R}$)	0.05	0.06	0.29	0.25	0.37	0.35	0.50	0.54
se($\mathbf{\Sigma}_3^{*R}$)	0.06	0.06	0.29	0.25	0.37	0.36	0.51	0.55
se($\mathbf{\Sigma}_3^{**R}$)	0.06	0.06	0.29	0.25	0.37	0.36	0.52	0.55
Proportional assignment								
Estimate	0.09	0.39	-0.12	-0.37	0.05	-0.74	0.17	-0.90
se($\mathbf{\Sigma}_3^{H}$)	0.06	0.06	0.35	0.27	0.44	0.38	0.59	0.53
se($\mathbf{\Sigma}_3^{*H}$)	0.07	0.06	0.35	0.27	0.44	0.38	0.59	0.54
se($\mathbf{\Sigma}_3^{**H}$)	0.07	0.06	0.35	0.27	0.44	0.38	0.59	0.53
se($\mathbf{\Sigma}_3^{R}$)	0.05	0.05	0.28	0.25	0.35	0.34	0.45	0.50
se($\mathbf{\Sigma}_3^{*R}$)	0.06	0.06	0.29	0.25	0.36	0.34	0.46	0.51
se($\mathbf{\Sigma}_3^{**R}$)	0.06	0.06	0.29	0.25	0.35	0.34	0.46	0.50
One-step								
Estimate	0.09	0.38	-0.11	-0.35	0.03	-0.72	0.13	-0.84
se($\mathbf{\Sigma}_1^{H}$)	0.05	0.05	0.28	0.24	0.36	0.34	0.51	0.48

Note: The H and R term refers to the choice of Σ_3. In all cases the Σ_2^H is used

then males.

The choice of standard error estimator is important when deciding about significance. The results show that using the Hessian based estimators for proportional assignment the effect of race on LC membership is non significant, while this effect is significant with all choices of standard error estimators for modal assignment and for one-step method. Once the robust standard error estimates are used, the effect of race is significant also with proportional assignment (for with and without uncertainty correction). The effect of sex is non significant and the effect of education is significant irrespective of the choice of standard error estimator.

Looking on the results obtained with modal assignment we see that the uncertainty uncorrected estimator, Σ_3 obtained lower estimates then the two correction methods, Σ_3^* and Σ_3^{**} for both choices of Σ_3, that is Hessian and robust (Σ_3^H and Σ_3^R). The two correction methods ($\Sigma_3^*, \Sigma_3^{**}$) obtain values that are close to each other. When comparing the Hessian and robust variance estimators to each other, we can see, that the former one obtains slightly lower estimates of the standard errors.

In case of proportional assignment all the estimates based on the Hessian matrix ($\Sigma_3^H, \Sigma_3^{H*}\Sigma_3^{H**}$) are overestimated as expected, while the estimates based on the robust variance estimator ($\Sigma_3^R, \Sigma_3^{R*}\Sigma_3^{R**}$)are lower, and close to the values obtained with modal assignment and the one-step method. Comparing the 3 options for uncertainty correction, similarly to modal assignment we can see that the uncorrected estimator (Σ_3) obtains lower estimates then the two corrections, irrespective of the choice for robust or Hessian based estimator. But for proportional assignment, the choice of the robust estimator makes the biggest difference. These results are in agreement with what was found in the simulation study.

6 Discussion

This paper contributes to the literature on three-step LCA by showing how correct standard error estimates of the third step model of the bias-adjusted three-step LCA can be obtained, by focusing on two main aspects.

On one hand we show the presence of an additional source of uncertainty, that is due to not accounting for the estimated nature of the fixed parameter values used in the last step. To solve this problem we show how the general theory of pseudo maximum likelihood estimation (Gong and Samaniego, 1981) can be applied to account for the additional uncertainty. Based on this general theory we propose two correction methods, one that accounts only for the uncertainty due to using fixed parameter values (denoted by Σ_3^*), and an extension of this that accounts also for same sample dependency of the parameters from the first and last step (denoted by Σ_3^{**}).

On the other hand we address a specific of LCA, namely that the choice of variance estimator of the free parameters depends on the choice of class assignment done in step two. For proportional assignment we suggested the use of the complex sampling variance estimator (denoted by Σ_3^R), as introduced by Michel, Hofstede and Steenkamp (1998) in the context of LCA, because this estimator can account for the multiple records per case. For modal assignment we argue for the Hessian based variance estimator.

We compared via a simulation study the performance of the different variance estimators. Based on the results of the simulation study we can conclude that in situations where the uncertainty about the fixed parameters is high an uncertainty correction should be used. In these situations the use of Σ_3^* suffices. This estimator is easier to obtain then Σ_3^{**}, and the difference between the two is negligible in the situations we examined. In situations where the uncertainty about the fixed estimates is low, that is the classification is good (the entropy R^2 is high) and/or the sample size is high the use of the correction methods is not needed. Based on the simulation conditions we can conclude that in the situations where the sample size is 2000 or higher, or the entropy is .90 or higher the uncertainty correction is not necessary.

Next, the results of the simulation study showed that the best variance estimator for proportional assignment is the complex sampling variance estimator, Σ_3^R. On the other hand, for the modal assignment the use of the observed Hessian based variance estimator(Σ_3^H) is recommended, because the robust variance estimator tends to overestimate the variance (Kauermann and Carroll, 2001). In the first step, irrespective of the assignment method used later on the use of the Hessian based variance estimator (Σ_1^H) is recommended.

Based on this study, the different variance estimators have been implemented in an experimental version of the standard latent class analysis software Latent Gold. This implementation makes the methods discussed in this paper directly available to applied researchers.

A limitation of our paper is that in the simulation study we restricted ourselves to situations where all model assumptions are met. There is a need for further research to see whether in situations where model assumptions are violated the same estimators should be chosen. For instance if in the first-step the model assumptions are violated the choice of the robust variance estimator might be recommended. Further research can also look into more complex models, whether the same conclusions would be reached.

References

Ahlquist, John S and Christian Breunig. 2012. “Model-based clustering and typologies in the social sciences.” *Political Analysis* 20:92–112.

Asparouhov, Tihomir and Bengt Muthén. 2012. “Auxiliary variables in mixture modeling: A 3-step approach using Mplus.” *Mplus Web Notes* 15.

Bakk, Zsuzsa, Fetene T. Tekle and Jeroen K. Vermunt. in press. “Estimating the Association Between Latent Class Membership and External variables Using Bias-adjusted Three-step Approaches.” *Sociological Methodology*.

Bandeen-Roche, Karen., Diana L. Miglioretti and Paul J. Zegger, Scott L. Rathouz. 1997. “Latent variable regression for multiple discrete outcomes.” *Journal of the American Statistical Association* 92:1375–86.

Bolck, Annabelle, Marcel Croon and Jacques Hagenaars. 2004. “Estimating Latent Structure Models with Categorical Variables: One-Step versus Three-Step Estimators.” *Political Analysis* 12:3–27.

Collins, Linda M and Stephanie T Lanza. 2010. *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*. Vol. 718 Wiley.

Dias, Jose G. and Jeroen K. Vermunt. 2008. "A bootstrap-based aggregate classifier for model-based clustering." *Computational Statistics* 23:643–59.

Feingold, Alan, Stacey S. Tiberio and Deborah M Capaldi. 2013. "New Approaches for Examining Associations with Latent Categorical Variables: Applications to Substance Abuse and Aggression." *Psychology of Addictive Behaviors* 1.

Gong, Gail and Francisco J Samaniego. 1981. "Pseudo maximum likelihood estimation: Theory and applications." *The Annals of Statistics*.

Goodman, Leo 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*.

Grimmer, Justin. 2013. "Appropriators not position takers: The distorting effects of electoral incentives on congressional representation." *American Journal of Political Science*.

Grimmer, Justin and Brandon M Stewart. 2013. "Text as data: The promise and pitfalls of automatic content analysis methods for political texts." *Political Analysis*.

Hagenaars, Jacques A. 1990. *Categorical longitudinal data- Loglinear analysis of panel, trend and cohort data*. Newbury Park, CA:Sage.

Hagenaars, Jacques A. 1993. *Loglinear models with latent varaibles*. Newbury Park, CA:Sage.

Kauermann, Göran and Raymond J. Carroll. 2001. "A Note on the Efficiency of Sandwich Covariance Matrix Estimation." *Journal of the American Statistical Association* 96:1387–1396.

McCutcheon, Allan L. 1985. "A latent class analysis of tolerance for nonconformity in the American public." *Public Opinion Quarterly* 49:474–488.

McCutcheon, Allan L. 1987. *Latent Class Analysis*. Newbury Park, CA:Sage.

Michel, Wedel., ter Frenkel Hofstede and Jan-Benedict E.M. Steenkamp. 1998. "Mixture model analysis of complex samples." *Journal of Classification*.

Murphy, Kevin M. and Robert H. Topel. 1985. "Estimation and Inference in Two-step Econometric Models." *Journal of Business and Economic Statistics* 3:88–97.

Mustillo, Thomas J. 2009. "Modeling new party performance: A conceptual and methodological approach for volatile party systems." *Political Analysis* 17:311–332.

Oberski, Daniel L. and Albert Satorra. 2013. "Measurement Error Models with Uncertainty about the Error Variance." *Structural Equation Modeling* 20.

Oehlert, Gary W. 1992. "A note on the delta method." *The American Statistician* 46:27–29.

Parke, William R. 1986. "Pseudo Maximum Likelihood Estimation: the Assymptotic Distribution." *The Annals of Statistics* 14:355–357.

Skrondal, Anders and Jouni Kuha. 2012. "Improved Regression Calibration." *Psychometrika* 77:649–669.

Venables, W. N., D. M. Smith and the R Core Team. 2013. "An Introduction to R. Notes on R: A Programming Environment for Data Analysis and Graphics Version 3.0.0.".
URL: *http://cran.r-project.org/doc/manuals/R-intro.pdf*

Vermunt, Jeroen K. 2010. "Latent Class Modeling with Covariates: Two Improved Three-Step Approaches." *Political Analysis* 18:450–469.

Vermunt, Jeroen K. and Jay Magidson. 2005. *Technical Guide for Latent GOLD 4.0: Basic and Advanced*. Belmont Massachusetts:Statistical Innovations Inc.

White, Halbert. 1982. "Maximum likelihood estimation of misspecified models." *Econometrica: Journal of the Econometric Society* 50:1–25.