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

Zsuzsa Bakk Daniel L. Oberski Jeroen K. Vermunt

Department of Methodology and Statistics

Tilburg University, The Netherlands

Address:

Room P 1.113, PO Box 90153, 5000 LE Tilburg

Phone: +31 13 466 2610

Email: z.bakk@tilburguniversity.edu

Funding

This work was supported by The Netherlands Organization for Scientific Research (NWO) [vici grant number 453-10-002].

Acknowledgements

Thanks are due to the anonymous reviewers and the editor, whose helpful comments improved the paper considerably.

Abstract

Latent class analysis is used in the political science literature in both substantive applications and as a tool to estimate measurement error. Many studies in the social and political sciences relate estimated class assignments from a latent class model to external variables. Though common, such a “three-step” procedure effectively ignores classification error in the class assignments; Vermunt (2010) showed that this leads to inconsistent parameter estimates and proposed a correction. Although this correction for bias is now implemented in standard software, inconsistency is not the only consequence of classification error. We demonstrate that the correction method introduces an additional source of variance in the estimates, so that standard errors and confidence intervals are overly optimistic when not taking this into account. We derive the asymptotic variance of the third-step estimates of interest, as well as several candidate corrected sample estimators of the standard errors. These corrected standard error estimators are evaluated using a Monte Carlo study and we provide practical advice to researchers as to which should be used so that valid inferences can be obtained when relating estimated class membership to external variables.

1 Introduction

Latent class analysis (LCA) is a tool used to classify objects for further analysis (Ahlquist and Breunig, 2012), with a wide range of applications in political science. 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 volatile party systems. Furthermore 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; and Grimmer (2013) related latent classes obtained from US Senators' press releases to their publicly expressed priorities. Further applications of LCA in political science include Feick (1989), Sniderman et al. (1989), Breen (2000), Hill and Kriesi (2001), Blaydes and Linzer (2008), Linzer (2011), Glasgow, Golder and Golder (2012), Ristei Gugiu and Centellas (2013), and Beissinger (2013). While most applications we refer to are substantive, LCA, and in general latent variable models can be used in a more instrumental manner as well, as a tool to estimate measurement error in observed variables (Fuller, 1987; Alwin, 2007; Fornell and Larcker, 1981; Rabe-Hesketh, Skrondal and Pickles, 2001; Oberski and Satorra, 2013).

As the above examples already suggest in most applications the interest lies not only in creating a latent classification, but also in relating this to external variables of interest. Usually this is done using a three-step procedure, even though a simultaneous estimation procedure is also available (Dayton and Macready, 1988; Hagenaars, 1990, 1993; Bandeen-Roche et al., 1997). The three-step approach proceeds as follows: in the first step, the latent class model is estimated; secondly units are classified into classes using some assignment mechanism based on the first step; and thirdly the newly created observed variable is related to external variables using standard methods such as (logistic) regression. Note that in the second step a classification error is introduced because the true class membership is unknown, unless there is perfect classification, and this error leads to biased parameter estimates in the third step (Vermunt, 2010; Bolck, Croon and Hagenaars, 2004).

Bias notwithstanding the three-step approach is still very popular in applied social and bio-medical research (Olino et al., 2010; McCutcheon, 1985; Clark and Besterfield-Sacre, 2009; Marsh et al., 2009; Loken, 2004; Chan and Goldthorpe, 2007). This popularity can be explained among other factors by the intuitive nature of the approach: researchers prefer to first establish a measurement model or a construct, and later regress the construct on potential predictors (Vermunt, 2010). The stepwise approach is preferred even more in situations where the classification needs to be related to dependent variables (distal outcomes). The reason for this preference is that if the dependent variables are added in a single step these variables would define the classification, whereas the intent is to explain them by the classification (Bakk, Tekle and Vermunt, 2013; Lanza, Tan and Bray, 2013), thus an unintended circularity would be created. In many situations the different steps are performed by different researchers, at different points in time. The stepwise approach can also be used in situations where the simultaneous estimation would be impossible, for instance when information about the classification error comes from a different sample.

Inspired by the widespread use of the three-step approach, Vermunt (2010) provided an improved three-step procedure in which the third step is amended by correcting for classification errors, thus removing the parameter bias. Bakk, Tekle and Vermunt (2013) and Asparouhov and Muthén (2012) tested Vermunt's approach via simulation studies by using models with distal outcome variables and latent transition analysis respectively, showing that in all these situations the bias-adjusted three-step approach performs well with regard to parameter bias reduction. Furthermore Feingold, Tiberio and Capaldi (2013) applied the corrected three-step approach to substance abuse data, and implementations of the method are available in standard latent class software Mplus Version 7.1 (Asparouhov and Muthén, 2012) and Latent GOLD Version 5.00 (Vermunt and Magidson, 2013).

As such the improved three-step procedures of Vermunt (2010) are easy to use due to their availability in mainstream software. However, as we show in this article, even after correcting for parameter bias an additional source of error remains, namely the three-step procedure causes additional variance in the estimates that should be accounted for. Depending on the assignment method used in step two, standard errors will be over- or underestimated (Vermunt, 2010; Bakk, Tekle and Vermunt, 2013) in the last step. This means that even though the parameter estimates are correct, statistical inferences are not. When underestimated standard errors are used the confidence intervals will be too narrow and significance tests overly optimistic, thus increasing the probability of Type I error. At the same time, using overestimated standard errors leads to loss of power. Considering the broad applications of three-step latent class modeling, this is an undesirable situation.

The problem of additional variance caused by using estimates from a previous step has been dealt with in the context of non-linear models (Carroll et al., 2006), and three-step structural equation modeling (Skrondal and Kuha, 2012; Oberski and Satorra, 2013), and econometric theory for two-stages least squares is already well-developed (Murphy and Topel, 1985). In this paper we apply the general theory of Gong and Samaniego (1981) to latent class modeling, noting similarities and differences with these other approaches.

In this article we introduce two correction methods that are based on the general theory of Gong and Samaniego (1981) and can account for the bias in the standard errors. We evaluate different possible estimators of the standard errors using Monte Carlo simulations showing how the optimal variance estimator depends on the class assignment method. We also provide advice which estimators to use in different situations in order to obtain correct inferences. Based on this study, the methods discussed have been made available to applied researchers in the syntax version of the software Latent GOLD 5.00 (Vermunt and Magidson, 2013).

Whereas most of this paper focuses on correct inferences using Vermunt's approach that conditions on the first step ML estimates (an approach that can be used in most practical situations), in the discussion we introduce the possibility of Bayesian inference, showing how the uncertainty about the first step parameters can be accounted for in the last step using multiple imputation. The Bayesian approach can be useful for instance in situations where model uncertainty is high, or sample size is low and there are strong priors available.

The structure of the paper is as follows: in Section 2 we introduce the bias-adjusted three-step latent class analysis. Next section 3 presents 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 revisits McCutcheon's (1985) analysis of how education and age groups differ in their tolerance. While the author initially used the uncorrected three-step approach, we show how inferences change using the corrections we propose. We conclude in Section 6, also showing directions for the Bayesian implementation of the methods we propose.

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; Hagenaars, 1990):

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 estimated classification W to the external variables of interest.

Note that the assigned scores on W obtained in the second step do not correspond exactly to the true values unless the classification is perfect. Thus, classification error is introduced. As a consequence the parameter estimates of the third-step model will be biased, since a model with variables with measurement error is estimated (Bolck et al., 2004; Hagenaars, 1990). However, a specific of this model is that the amount of error introduced in step two is known. Thus, the step three model can be augmented to account for this known classification error (Bolck et al., 2004; Vermunt, 2010). In the following we introduce in detail each of the steps, explaining how the step three model is corrected for.

Special attention is given to the possible variance estimators. In each step a choice can be made between Hessian or robust variance estimator, nevertheless it is not clear which is better. More importantly in the third step the variance estimators should also account for the additional variance due to the correction method implemented. We propose two correction methods, and later in the simulation study we cross the choice of Hessian or robust estimator with the choices of correction methods to give practical advice on which variance estimator to use.

Similar models, in which the amount of error in the proxy is known or the true value is approximated via multiple proxies, are available in political science literature Blackwell, Honaker and King (2012) and in the literature on measurement error correction via latent variable models Alwin (2007); Fornell and Larcker (1981); Skrondal and Kuha (2012); Oberski and Satorra (2013).

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). Where by indicator variables we understand, the observed variables used to define the LC model. Given a sample of n units, the observations $\mathbf{Y}_i$ are modelled as arising from T unobserved (latent) classes X,

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

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.

Note that in most applications the number of latent classes is not known a priori, but can be selected based on a set of modification indices (AIC, BIC). While selecting the right number of classes is outside the focus of this paper, we recommend for those interested in this problem to refer to Nylund, Asparouhov and Muthén (2007), Van der Heijden, 't Hart and Dessens (1997), or Sclove (1987).

Further, 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,

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

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 case wise 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.

In situations where the misspecification is strong using robust standard errors does not suffice. As King and Roberts (2012) highlights in these situations instead of relying on robust standard errors it is recommended to check where the misspecification is located, and correct for that. Some useful tools to check for misspecification are: the BVR statistics that checks whether there is residual association between two indicators after controlling for latent class membership (Vermunt and Magidson, 2013, p.72-73), or the EPC statistics, that shows how much the model parameters would change if one parameter was freed (Oberski and Vermunt, 2013). These statistics are especially useful to test whether there is a direct effect between an indicator and external variable or if there is residual association left between the indicators after controlling for class membership. If such effects exist they should be modelled in the step one model.

For simplicity of exposition in the following we restrict ourselves to models where all model assumptions hold, that is the conditional independence assumption of the indicators holds, and there are no direct effects of external variables on the indicators. Furthermore we assume that the number of classes is known. [1]

2.2 Second step: assignment of units to classes

After estimating 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

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

Sample estimates of the posterior probabilities $P(X_i = t|\mathbf{Y}_i)$ 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 $\hat{\boldsymbol{\theta}}_1$. These estimates can be used in different ways to create an estimated class membership variable W (Vermunt, 2010). We introduce the two most

[1] In situations where there is model uncertainty in step one a Bayesian approach might be used, as introduced in the discussion. Note that uncertainty about the number of classes cannot easily be handled with the Bayesian approach, only uncertainty about direct effects.

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 rule (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 $P(X_i = t|\mathbf{Y_i}) > P(X_i = t^{'}|\mathbf{Y_i})$ 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 $P(X_i = t|\mathbf{Y}_i)$. 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, 2013),

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

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 per assignment method. For example, $P(W_i = s|\mathbf{Y}_i)$ is either 0 or 1 using modal assignment, and with proportional assignment $P(W_i = s|\mathbf{Y}_i) = P(X_i = s|\mathbf{Y}_i)$.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 class 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:

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

Details of the derivation are available in Bolck et al. (2004). Equation 10 can be seen as a latent class model with W as a single indicator that is fixed to the“known” classification error probabilities $P(W = s|X = t)$, as defined in Equation 8 Vermunt (2010). 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 (2013) 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, 2013).

In the following we introduce two correction methods to account for this additional uncertainty. We also highlight a special problem of proportional assignment that needs to be solved regardless of the choice made for correction for uncertainty in the variance estimator.

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

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

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

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

where $\mathbf{\Sigma}_3$ 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 (2013) found that using the third-step Hessian matrix to obtain an estimator of $\mathbf{\Sigma}_3$, 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 (Wedel, Ter Hofstede and Steenkamp, 1998), which will be denoted $\mathbf{\Sigma}_3^R$. Using this 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, $\mathbf{R} = (\partial L_3 / \partial \boldsymbol{\theta}_3)'(\partial L_1 / \partial \boldsymbol{\theta}_1)$. 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, $\mathbf{\Sigma}^{**} \stackrel{a}{=} \mathbf{\Sigma}^{*}$, although they may not be equal in small samples. In small samples it is possible that $\mathbf{\Sigma}^{*}$ will overestimate the standard errors of the third-step estimates, although this overestimation should decrease as sample size increases; on the other hand, the calculation of the extra terms in $\mathbf{\Sigma}^{**}$ may add considerable effort and instability to the standard errors.

Whether $\mathbf{\Sigma}^{*}$ or $\mathbf{\Sigma}^{**}$ 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 exist. 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 $\mathbf{\Sigma}_2$ and $\mathbf{\Sigma}_3$: observed Hessian based or robust) with the options for correcting for uncertainty ($\mathbf{\Sigma}_3$ - uncorrected, $\mathbf{\Sigma}_3^{*}$ first order and $\mathbf{\Sigma}_3^{**}$ second order correction) for both modal and proportional assignment. In the following table we summarize the different choices of the variance estimators compared.[2]

As used in Table 1 the 1^{st} order correction, $\mathbf{\Sigma}_3^{*}$ is defined in equation 17 and the 2^{nd} order correction, $\mathbf{\Sigma}_3^{**}$ in 18, and $\mathbf{\Sigma}_3$ is the variance of the free parameters ignoring the additional uncertainty attributable to the fixed parameter values. In reporting the simulation 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 (Wedel, Ter Hofstede and Steenkamp, 1998), and for modal assignment for the sandwich estima-

[2]The simulation set up is available in the dataverse replication material of this article: Study Global Id: doi:10.7910/DVN/24497 (v1)

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 ($\mathbf{\Sigma}_3^R$)

tor 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. Where we use the first subscript for Z, and the second subscript for X, as such for example β_{13} stands for the effect of Z_1 on the third class. Te effect of Z_2 on X is set to 1 (β_{22}) and 0 (β_{23}), 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, 2013) and R (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 estimates obtained using 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, 2013; Vermunt, 2010), and show that the three- step approach can be used without loss of efficiency or parameter bias.

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

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 ($\mathbf{\Sigma}_3^*$ or $\mathbf{\Sigma}_3^{**}$) improves the results for both Hessian based and robust estimator. Comparing the first and second order corrections ($\mathbf{\Sigma}_3^*$, $\mathbf{\Sigma}_3^{**}$) 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 ($\mathbf{\Sigma}_3, \mathbf{\Sigma}_3^*, \mathbf{\Sigma}_3^{**}$). This can be seen from both the se/sd (which is higher then 1 for all three estimators) and

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

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 for all the variance estimators the ratio of the standard error to the standard deviation (se/sd) is between 1.00-1.02, and the coverage rate is 96 %, results that show that all standard error estimators perform well in this condition. 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

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

condition are presented. For modal assignment the uncertainty uncorrected standard error 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 assign-

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

ment 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.[3] 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 Example application using the suggested corrections: cohort and education effects on tolerance

We now show how correcting for parameter bias in the three-step latent class analysis makes a difference for substantive conclusions. For this purpose we re-analyze the often-cited example of latent class analysis in political science, McCutcheon (1985)'s assessment of how age and education groups differ in their tolerance towards out-groups. In addition, we illustrate how the different choices of standard error discussed above can affect results.

The question of who is more intolerant originated with Stouffer (1955). His analysis, conducted at the height of the McCarthy era, focused on citizens' tolerance for communists: should they be allowed basic democratic rights to free speech according to the public? McCutcheon (1985) re-assessed this question by including other groups in the questionnaire besides communists, namely atheists, homosexuals, mil-

[3] This tendency can be seen in more detail in the Appendix in Tables A1 and A2, which show that once the uncertainty decreases (either by a stronger effect of classes on indicators or higher sample size) the effect of the corrections is lower, but when the uncertainty is high the corrections make a big difference.

itarists, and racists. He then used a latent class model to integrate respondents' tolerance for these different groups into a single categorical latent variable that represents each person's "tolerance for nonconformity". After estimating this model (step 1) and assigning a "tolerance" classification to each respondent (step 2), McCutcheon then regressed the categorical "tolerance" assignment on age cohort and educational attainment (step 3). This commonly cited example of a latent class analysis in political science is therefore in fact a bias-uncorrected three-step analysis.

Here we examine how McCutcheon (1985)'s conclusions might change when the necessary (e.g. Roeder, Lynch and Nagin, 1999; Bolck, Croon and Hagenaars, 2004) bias corrections (Vermunt, 2010) are applied to the third step, predicting "tolerance" from age cohort and education. We also show how the standard errors of this regression differ over the various choices of standard error estimator described in the preceding sections.

The original data are obtained from the 1976 and 1977 General Social Survey (GSS), which are publicly available[4]. Each of the 2689 respondents (nr. of respondents obtained after listwise deletion was applied) answered the following questions on communists, atheists, homosexuals, militarists, or racists:

- "Suppose this _____ wanted to make a speech in your community. Should he be allowed to speak?" (**Yes**/No)
- "Should such a person be allowed to teach in a college or university, or not?" (**Yes**/No)
- "Suppose he wrote a book which is in your public library. Somebody in your community suggests that the the book should be removed from the library. Would you favor removing it or not?" (Yes/**No**)

The bold-faced answers are those indicating tolerance. McCutcheon coded a re-

[4] https://www.icpsr.umich.edu/icpsrweb/ICPSR/series/28/studies/7398?archive=ICPSR&sortBy=7, also available in the dataverse replication material of this article: Study Global Id: doi:10.7910/DVN/24497 (v1)

Table 6: Fit of different latent class models to five indicators of tolerance for nonconformity from the 1976/77 General Social Survey (N=2689). Shown are the likelihood ratio (L^2), degrees of freedom (df), Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC), and the largest bivariate residual (BVR) in the pairwise cross-table of the five indicators.

	L^2	df	p	BIC	AIC	max(BVR)
Independence model	4695.8	26	0.00	4490.5	4643.8	1048.5
2-class model	240.8	20	0.00	82.9	200.8	13.2
3-class model	48.7	14	0.00	-61.7	20.7	6.0
4-class model	6.5	8	0.59	-56.6	-9.4	0.11
5-class model	2.9	2	0.24	-12.9	-1.1	0.08

spondent as “tolerant” towards a group if all three of these bold-faced answers were given, and “intolerant” otherwise. This yields five binary indicators of general tolerance (one for each group).

The first step is to fit local independence models with a successively increasing number of latent classes. Resulting model fit statistics are shown in Table 6. This Table shows that the AIC selects the four-class model, while the BIC selects the three-class model. Looking at the absolute model fit of the four-class model, however, it is clear that both the likelihood ratio as well as the largest bivariate residual (BVR; see Vermunt and Magidson 2013) indicate residual dependencies between the indicators in the three-class model. We therefore follow McCutcheon (1985) and select the four-class model, which fits the data well. None of the residual dependencies between the observed variables are substantively large or statistically significant in the four-class model, thus we can reasonable assume that there is no residual variance left between the indicators.

Estimates of class sizes and conditional probabilities obtained from the four-class model are shown in Table 7. The first row of this table indicates the labels given to the four classes. These labels are based on the pattern of estimated conditional probabilities given on the subsequent rows. For example, the first class, to which 56% ($\pm 2\%$) of respondents are estimated to belong, exhibits low probabilities of toler-

Table 7: Parameter estimates (standard errors) for the four-class model: class sizes and conditional probabilities to give all-tolerant answers given the latent class.

	"Intolerant"		"Tolerant"		"Intolerant of right"		"Intolerant of left"	
Class size	0.56	(0.02)	0.23	(0.01)	0.11	(0.03)	0.10	(0.03)
Tolerance for...								
Atheists	0.03	(0.01)	0.98	(0.01)	0.41	(0.06)	0.61	(0.07)
Communists	0.04	(0.01)	0.95	(0.02)	0.59	(0.11)	0.27	(0.07)
Militarists	0.05	(0.01)	0.92	(0.02)	0.34	(0.05)	0.38	(0.06)
Racists	0.08	(0.01)	0.90	(0.02)	0.02	(0.06)	0.81	(0.20)
Homosexuals	0.13	(0.01)	0.96	(0.01)	0.72	(0.07)	0.56	(0.06)

ance for all groups, ranging between 0.03 for atheists to 0.13 for homosexuals (both ± 0.01). Therefore this latent class was labeled "intolerant". There are also classes of those respondents who are tolerant towards some groups and not others. Since these preferences appear to correspond to political ideology, McCutcheon labeled the classes "intolerant of right" and "intolerant of left" respectively. It should be noted, however, that this ideological intolerance is not symmetric: those who are intolerant of "right-wing" groups such as racists and militarists are more extreme in their opinion than those who are intolerant of "leftist" groups such as atheists or communists. The difference between these classes in their tolerance for homosexuals is not statistically significant ($z = -1.52$, $p = 0.064$).

While the latent class analysis of these five indicators (step 1) is interesting in itself, to Stouffer (1955) and McCutcheon (1985) the main substantive question was how age cohorts and educational groups differ in their overall tolerance. For this reason, each respondent was assigned to one of the four estimated classes using proportional assignment based on the latent class model (step 2). This assigned class variable is highly convenient for further analysis: the analyst performing the latent class analysis may be separate from the researcher investigating substantive questions using the result. Thus, the researcher interested in the effect of cohort and education on tolerance need not have the same expertise as the latent class analyst.

Furthermore, the definition of the latent class assignment has not been affected by the cohort and education variables, preventing circularity in the final results.

Following McCutcheon (1985), educational attainment was coded into three categories: those with fewer than twelve grades (1), those who completed high school (2), and those with more than twelve years of formal education (3). Birth cohort was coded into four categories: those born in or before 1914 (4), those born between 1915 and 1933 (3), those born between 1934 and 1951 (2), and those born after 1951 (1). We ran a multinomial regression of assigned tolerance on these covariates, corrected for misclassification error in the assignment and using the first categories of each variable as a reference category. The interaction effect of education $\times$ cohort turned out to be small and not statistically significant (Wald = 12.2 on 18 df, $p = 0.84$), and we therefore decided to exclude the interaction from further analysis. The resulting main effect estimates are shown as points with 95% confidence intervals in Figure 1.

Figure 1 summarizes the 15 multinomial logit coefficients and their standard errors from the uncorrected analysis (black dots) performed by McCutcheon (1985) and the corrected analysis (gray points). The effect size estimates show that the more educated and younger a person is, the more likely they are "tolerant". Looking at Figure 1 from bottom to top reveals a monotonic increase of logits with these two covariates. This applies to a lesser degree to the "intolerant to right" group, and to the "intolerant to left" group to an even lesser degree. This ordering in effect size from the rightmost to leftmost panel in Figure 1 is probably due to ideological differences between age and education groups: the younger and more educated were more likely to prefer the political left.

For comparisons between corrected and uncorrected analyses, the 15 coefficients from the corrected three-step analysis are shown below each estimate (gray points). It can be seen that there is a substantial difference between the corrected and the uncorrected estimates. To bring this point into perspective, Figure 2 shows the ratios of corrected to uncorrected point estimates. The most extreme case is the logit coef-

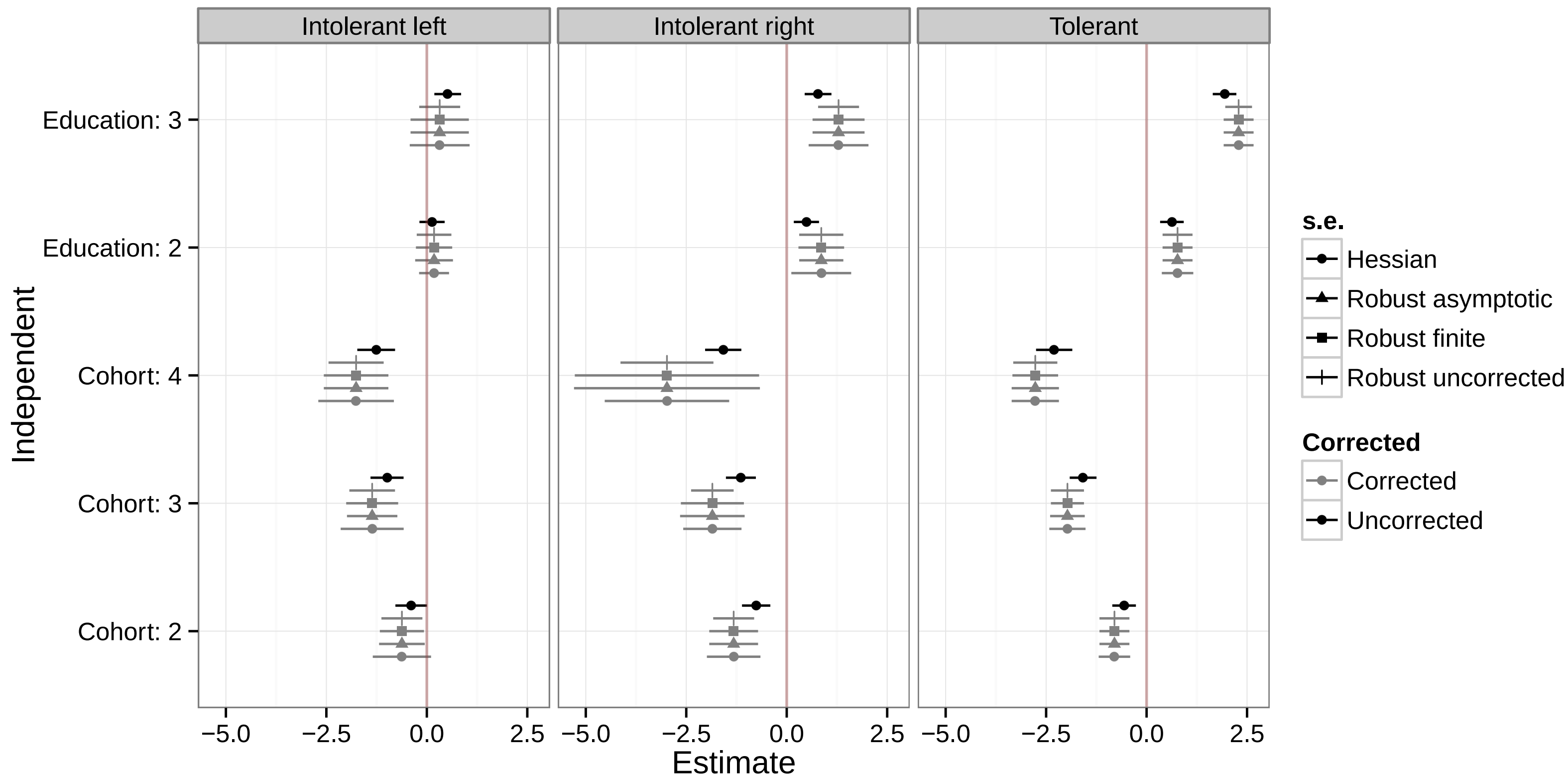

Figure 1: Multinomial effect size estimates of covariates on the tolerance classes from the third-step analysis. Shown are the dummy-coded effects of two education categories and three age cohorts on the four latent classes, using the first category as a reference. The uncorrected point estimate and 95% confidence interval is shown, below which the corrected point estimates with confidence intervals from four types of standard errors is given.

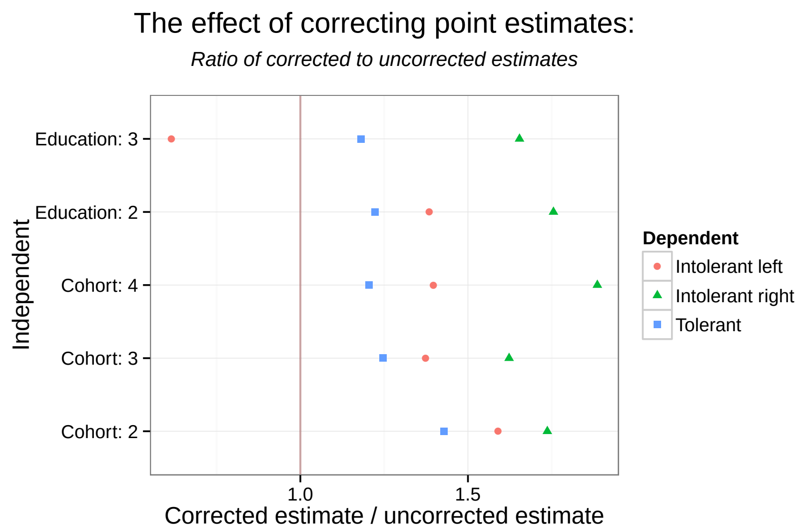

Figure 2: The relative size of the corrected point estimates compared with the uncorrected. Unity means the two estimates are the same, while 2 means the corrected point estimate is twice as large as the uncorrected estimate. The effects for different classes are indicated with different point shapes, given in the legend.

ficient of the oldest cohort on being in the "intolerant to right" class, which increases almost twofold. Even the smallest correction entails a 20% larger coefficient, however. This emphasizes the practical significance of correcting for classification error in the class assignment: substantial bias will otherwise occur.

While the first-step sample size of 2689 is relatively large, the entropy R^2 of the "tolerance" latent classification is 0.71, falling short of the 0.90 our simulations identified as an indicator of "small" uncertainty about the classification error. The standard errors of this analysis may therefore benefit from correction for this uncertainty.

Does the correction to standard errors introduced in this paper make a difference for the results? Figure 1 shows that it does. First of all, qualitative differences can be observed for the effect of being in the younger cohort (2) on being "intolerant to left": this cohort is no longer deemed to differ significantly from the youngest cohort (1) when the correction is applied.

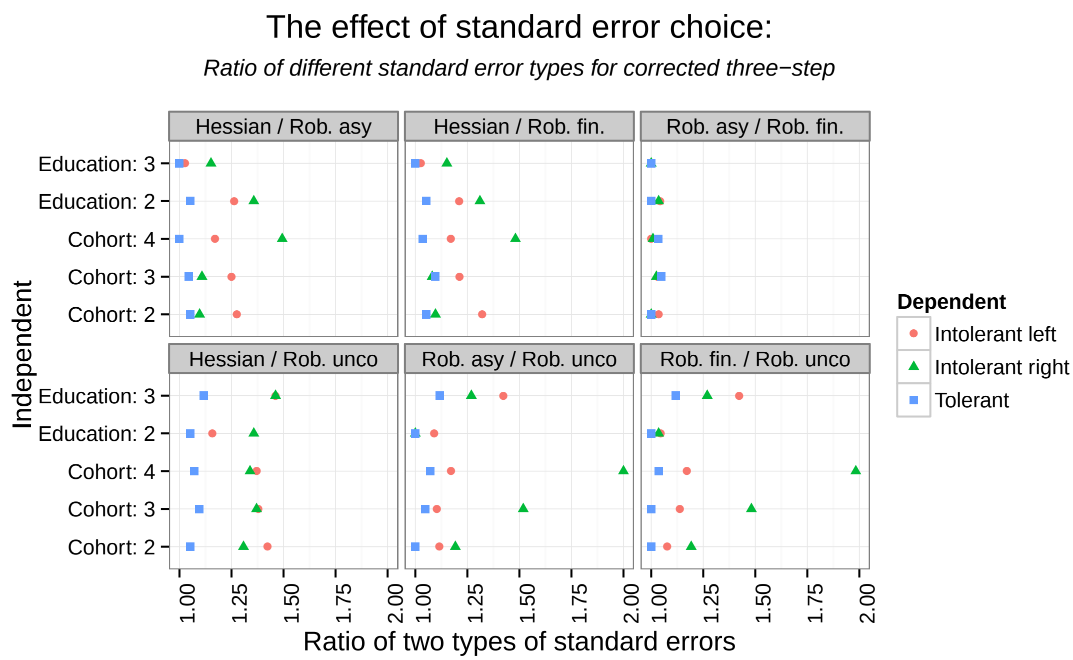

Figure 3: Comparison between the four different types of standard errors for the corrected procedure. Shown are the ratios of the larger standard error (se) to the smaller se for the six combinations shown in the panel titles.

Second, Figure 1 exhibits large quantitative differences in standard errors, particularly for the effects of being in the oldest cohort (4). Figure 3 shows the sizes of these standard errors relative to each other. For instance, the finite-sample robust standard error for the oldest cohort's effect on being "intolerant to right" is twice as large as the standard error without the correction discussed above (triangle in the bottom-right graph). Since the relative standard error must be squared to obtain a "misspecification effect" (Skinner, Holt and Smith, 1989), this means that if the correction introduced here is ignored, the researcher will claim that the sample is four times more informative than it is. This clearly demonstrates the relevance of the corrections to standard errors introduced above.

6 Discussion and conclusion

Social scientists often aim to study the relationship of an unobserved classification with external variables. The "three-step" approach is common: a latent class model is fit (step 1), units are assigned to estimated classes (step 2), and the relationships of interest are studied using the assigned classes, for instance by multinomial regression (step 3). Three-step analysis potentially has several advantages over the "one-step" full-information maximum likelihood approach (see Vermunt, 2010, p. 451). Despite being common and attractive, this approach is also inconsistent—a problem solved by the bias-corrected three-step approach introduced by Bolck, Croon and Hagenaars (2004) and Vermunt (2010) in this journal (see also Bakk, Tekle and Vermunt, 2013).

Correct inferences about the relationships of interest, however, require not only consistent point estimates but also correct standard errors. In this article we show both analytically and by simulation that correct standard errors from the third step must incorporate the uncertainty about the classification error. Standard software allowing for bias-corrected three-step analysis such as Mplus (Muthén and Muthén, 1998-2012) and Latent GOLD (Vermunt and Magidson, 2013) did not incorporate this uncertainty, leading to underestimated standard errors and confidence intervals that are too narrow. We therefore provide in this paper the correct standard errors allowing for appropriate inferences, based on classic likelihood theory (e.g. Gong and Samaniego, 1981). As a result of our study, the different standard error estimators discussed have been implemented in the standard latent class analysis software Latent GOLD 5.00 (Vermunt and Magidson, 2013), making the methods developed here directly available to applied researchers.

Moreover, we evaluate eight possible types of standard errors. Although these standard errors are asymptotically equivalent under model correctness, they may yield different results in finite samples. A Monte Carlo simulation study compared them and found that the correction to standard errors introduced here can make a large difference when uncertainty about the first-step parameters is substantial. On the other hand, when the uncertainty about fixed estimates is low, the standard error corrections are not needed. Low uncertainty about the classification error will occur with large first-step sample sizes and high entropy R^2 (high class separation). No substantial differences between inferences based on corrected versus uncorrected standard errors were found with first-step sample sizes above 2000 combined with entropy $R^2 > 0.90$. We also noted little difference between an asymptotic and finite-sample version of the corrected estimator. The asymptotic corrected standard error estimator (Oberski and Satorra, 2013), which is considerably easier to compute, is therefore recommended. Finally, we reproduced the finding of Vermunt (2010) that proportional assignment requires robust standard errors to account for the replication of cases.

Considering these findings, bias correction and the choice of standard errors can make a difference for substantive conclusions. reanalysis of an example bias-uncorrected three-step analysis from the political science literature (McCutcheon, 1985) clearly demonstrated this effect. The logistic regression coefficients that give the strength of the relationship between being in the “intolerant to the right” class and all age and education categories are between one-and-a-half and twice as large as their uncorrected counterparts, for instance. The correction for classification error need not always increase estimated relationships: one of the coefficients of interest is much lower after correction. “Qualitative” differences also occur after correcting both the point estimates and the standard errors: the younger cohort’s logistic coefficient for “intolerant to left” is not statistically significant in the uncorrected analysis, but is so after correction. Conversely, the highly educated group’s coefficient is statistically significant in the uncorrected analysis, but not after correction. This demonstrates the importance of the corrections, both for point estimates and standard errors.

A limitation of our study is that we restricted ourselves to situations where the separation between classes is relatively good (entropy R^2= 60 or higher). We did so because previous research showed that in situations where the entropy is lower than this, the step three methods obtain biased estimates. This is due to the fact that in step one the classification error is underestimated, and thus over-optimistic correction terms are used (Vermunt, 2010, Bakk et al., 2013).

A further limitation of our study is that we assumed an (approximately) correct model can be found. That is, we assume that mostly sampling variance drives the first-step model uncertainty. For this reason, model checking in the first step is essential. A possible alternative approach would be to obtain point and uncertainty estimates under model uncertainty, after which these may be propagated to the third step as described above.

A completely different approach is the Bayesian multiple imputation framework (Rubin, 1987). In this framework, the first step is to formulate a Bayesian latent class model, the second to obtain M multiple draws from the latent class distribution, and the third to estimate M regression models, averaging the M parameter estimates and using the rules described by Schafer (1997) to correctly obtain standard errors. The idea of using Bayesian data augmentation to estimate the conditional distribution of a latent variable X given predictors Z was introduced by Tanner and Wong (1987). Applications of this idea can be found in the "plausible values" literature for continuous latent variables (e.g. Mislevy, 1988), as well as the method of "pseudo-class draws" (Bandeen-Roche et al., 1997; Wang, Brown and Bandeen-Roche, 2005; Asparouhov and Muthén, 2012).

However, the same inconsistency problems that plague the uncorrected three-step method also affect the Bayesian multiple imputation approach. The key difference between, three-step analysis, plausible values, and pseudo-class draws on the one hand,and the Bayesian augmentation literature, on the other is the inclusion of the covariates in the first-step model. As shown by Tanner and Wong (1987, pp. 530–1), the multiple imputations of the latent variable X must be generated from $p(X|Y, Z)$, not just $p(X|Y)$ as is done in the three-step, plausible values, and pseudo-class draws procedures. Leaving out the predictor variables Z from the first-step imputations will therefore cause the same inconsistency as is present in likelihood-based uncorrected three-step analysis. This was also demonstrated by the simulations of Asparouhov and Muthén (2012, Table 1).

However, including Z in the first-step analysis partially defeats the purpose of the three-step procedure. Moreover, researchers performing complicated latent variable models to publish imputations for the broader research community (e.g. König, Marbach and Osnabrügge, 2013) cannot possibly foresee all predictor variables Z that might someday be of interest.

The problem of inconsistency in the Bayesian multiple imputation approach caused by ignoring Z in the imputation model can in principle be solved by performing the bias correction described above in each imputation. The combined corrected point estimates will then be consistent for $p(X|Z)$. Afterwards, combining the resulting multiple corrected estimates will yield correct standard errors (Schafer, 1997). This method could substantially improve the quality of inferences from recent efforts in political science to publish multiple imputations of latent variables for further analysis such as Democracy (Treier and Jackman, 2008) or party positions (König, Marbach and Osnabrügge, 2013). It should be noted that the proposed correction method is not limited to Bayesian multiple imputation, but can be used in any situation in which an integration or missing data problem is solved by simulation and is based on a (step-one) model that excludes some of the relevant variables. The resulting corrected "pseudo-class draws" or "plausible values" analysis is an interesting and potentially useful application of the presented three-step approach that warrants future study.

References

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

Alwin, Duane, F. 2007. *Margins of error. A study of reliability in survey measurement*. New York, NY Willey.

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. 2013. "Estimating the association between latent class membership and external variables using bias-adjusted three-step approaches." *Sociological Methodology* 43:272–311.

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

Beissinger, Mark R. 2013. "The Semblance of Democratic Revolution: Coalitions in Ukraine's Orange Revolution." *American Political Science Review* 107:574–592.

Blackwell, Matthew, James Honaker and Gary King. 2012. "Multiple Overimputation: a unified approach to measurement error in missing data.".
URL: *http://gking.harvard.edu/files/gking/files/measure.pdf*

Blaydes, Lisa and Drew A. Linzer. 2008. "The political economy of women's support for fundamentalist Islam." *World Politics* 60:579–609.

Bolck, Annabelle, Marcel Croon and Jacques A. Hagenaars. 2004. "Estimating latent structure models with categorical variables: One-step versus three-step estimators." *Political Analysis* 12:3–27.

Breen, Richard. 2000. "Why is support for extreme parties underestimated by surveys? A latent class analysis." *British Journal of Political Science* 30:375–82.

Carroll, Raymond J., David Ruppert, Leonard A. Stefanski and Ciprian Crainiceanu. 2006. *Measurement error in nonlinear models: A modern perspective, Second Edition*. New York: Chapman and Hall/CRC.

Chan, Tak Wing and John H. Goldthorpe. 2007. "Social stratification and cultural consumption: music in England." *European Sociological Review* 23:1–19.

Clark, Renee M. and Mary E. Besterfield-Sacre. 2009. "A new approach to hazardous materials transportation risk analysis: decision modeling to identify critical variables." *Risk Analysis* 29:344–354.

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

Dayton, C. Mitchel and George B. Macready. 1988. "Concomitant-variable latent class models." *Journal of the American Statistical Association* 83:173–178.

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

Feick, Lawrence F. 1989. "Latent class analysis of survey questions that include don't know responses." *Public Opinion Quarterly* 53:525–47.

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*.
URL: *http://doi.apa.org/getdoi.cfm?doi=10.1037/a0031487*

Fornell, Claes and David Larcker. 1981. "Evaluating structural equation models with unobservable variables and measurement error." *Journal of Marketing Research* 18:39–50.

Fuller, Wayne, A. 1987. *Measurement error models*. New York, NY Willey.

Glasgow, G., M. Golder and S. N. Golder. 2012. "New Empirical Strategies for the Study of Parliamentary Government Formation." *Political Analysis* 20:248–270.

Gong, Gail and Francisco J. Samaniego. 1981. "Pseudo maximum likelihood estimation: Theory and applications." *The Annals of Statistics* 9:861–869.

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* 79:1179–1259.

Grimmer, Justin. 2013. "Appropriators not position takers: The distorting effects of electoral incentives on congressional representation." *American Journal of Political Science* 57:624–642.

Grimmer, Justin and Brandon M. Stewart. 2013. "Text as data: The promise and pitfalls of automatic content analysis methods for political texts." *Political Analysis* 21:267–297.

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 variables*. Newbury Park, CA:Sage.

Hill, Jennifer L. and Hanspeter Kriesi. 2001. "Classification by opinion-changing behavior: A mixture model approach." *Political Analysis* 9:301–324.

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.

King, Gary and Margaret Roberts. 2012. "How robust standard errors expose methodological problems they do not fix.".
URL: *http://gking.harvard.edu/files/gking/files/robust.pdf*

König, Thomas, Moritz Marbach and Moritz Osnabrügge. 2013. "Estimating Party Positions across Countries and TimeA Dynamic Latent Variable Model for Manifesto Data." *Political Analysis* 21:468–491.

Lanza, Stephanie T., Xianmin Tan and C. Bethany Bray. 2013. "Latent class analysis with distal outcomes: A flexible model-based approach." *Structural Equation Modeling* 20:1:1–26.

Linzer, Drew A. 2011. "Reliable inference in highly stratified contingency tables: using latent class models as density estimators." *Political Analysis* 19:173–187.

Loken, Eric. 2004. "Using latent class analysis to model temperament types." *Multivariate Behavioral Research* 39:625–652.

Marsh, Herbert W., Oliver Ludtke, Ulrich Trautwein and Alexandre J.S. Morin. 2009. "Classical latent profile analysis of academic self-concept dimensions: Synergy of person- and variable-centered approaches to theoretical models of self-concept." *Structural Equation Modeling* 16:191–225.

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.

Mislevy, R.J. 1988. Randomization-based inferences about latent variables from complex samples. Technical Report. Educational Testing Service.

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.

Muthén, L. K. and B. O. Muthén. 1998-2012. *Mplus Users Guide*. 7 ed. Los Angeles, CA: .

Nylund, Karen L., Tihomir. Asparouhov and Bengt Muthén. 2007. “Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study.” *Structural Equation Modeling* 14:535–569.

Oberski, Daniel and Jeroen K. Vermunt. 2013. “A Model-Based Approach to Goodness-of-Fit Evaluation in Item Response Theory.” *Measurement: Interdisciplinary Research & Perspectives* 11:117–122.

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.

Olino, Thomas M., Daniel N. Klein, Peter M. Lewinsohn, Paul Rohde and John R. Seeley. 2010. “Latent trajectory classes of depressive and anxiety disorders from adolescence to adulthood: Descriptions of classes and associations with risk factors.” *Comprehensive Psychiatry* 51:224–235.

Parke, William R. 1986. "Pseudo maximum likelihood estimation: the assymptotic distribution." *The Annals of Statistics* 14:355–357.

R Core Team. 2013. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0. **URL:** *http://www.R-project.org/*

Rabe-Hesketh, Sophia, Anders Skrondal and Andrew Pickles. 2001. "Maximum likelihood estimation of generalized linear models with covariate measurement error." *The Stata Journal*.

Ristei Gugiu, M. and M. Centellas. 2013. "The Democracy Cluster Classification Index." *Political Analysis* 21:334–349.

Roeder, Kathryn, Kevin G Lynch and Daniel S Nagin. 1999. "Modeling uncertainty in latent class membership: A case study in criminology." *Journal of the American Statistical Association* 94:766–776.

Rubin, Donald B. 1987. *Multiple imputation for nonresponse in survey research*. New York, NY Willey.

Schafer, J.L. 1997. *Analysis of Incomplete Multivariate Data*. Chapman & Hall/CRC.

Sclove, Stanley L. 1987. "Application of model-selection criteria to some problems in multivariate analysis." *Psychometrika* 52:333–343.

Skinner, C.J., D Holt and T.M.F. Smith. 1989. *Analysis of Complex Surveys*. New York, Wiley.

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

Sniderman, Paul M., Philip E. Tetlock, James M. Glaser, Donald Philip Green and Michael Hout. 1989. "Principled Tolerance and the American Mass Public." *British Journal of Political Science* 19:25.

Stouffer, Samuel Andrew. 1955. *Communism, conformity, and civil liberties: A cross-section of the nation speaks its mind*. Transaction Books.

Tanner, Martin A and Wing Hung Wong. 1987. "The calculation of posterior distributions by data augmentation." *Journal of the American statistical Association* 82:528–540.

Treier, Shawn and Simon Jackman. 2008. "Democracy as a latent variable." *American Journal of Political Science* 52:201–217.

Van der Heijden, Peter, Harm 't Hart and Jos Dessens. 1997. "A parametric bootstrap procedure to perform statistical tests in a LCA of anti-social behaviour." In *Applications of latent trait and latent class models in the social sciences.*, ed. J. Rost and R. Langeheine. New York, NY: Waxmann pp. 196–208.

Vermunt, Jeroen K. 2010. "Latent class Mmodeling with covariates: Two improved three-step approaches." *Political Analysis* 18:450–469.

Vermunt, Jeroen K. and Jay Magidson. 2013. *Technical guide for Latent GOLD 5.0: Basic and advanced*.

Wang, Chen-Pin, Hendriks C. Brown and Karen Bandeen-Roche. 2005. "Residual diagnostics for growth mixture models: Examining the impact of a preventive intervention on multiple trajectories of aggressive behavior." *Journal of the American Statistical Association* 100:1054–10.

Wedel, Michel, Frenkel Ter Hofstede and Jan-Benedict E.M. Steenkamp. 1998. "Mixture model analysis of complex samples." *Journal of Classification* 15:225–244.

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

Appendices

Table A1: Comparison of the different variance estimators for low separation,for the 3 sample sizes separatly for one parameter, β_{13} for modal and proportional assignment

Final	Components		500			1000			2000		
	2^{nd} step	3^{rd} step	se	se/sd	coverage	se	se/sd	coverage	se	se/sd	coverage
					Modal						
$\mathbf{\Sigma}_3$		$\mathbf{\Sigma}_3^H$	0.17	0.83	0.90	0.12	0.97	0.94	0.08	1.03	0.96
$\mathbf{\Sigma}_3^*$	$\mathbf{\Sigma}_2^H$	$\mathbf{\Sigma}_3^H$	0.20	0.98	0.95	0.13	1.06	0.97	0.09	1.12	0.97
$\mathbf{\Sigma}_3^{**}$	$\mathbf{\Sigma}_2^H$	$\mathbf{\Sigma}_3^H$	0.20	0.99	0.95	0.13	1.08	0.97	0.09	1.14	0.97
$\mathbf{\Sigma}_3$		$\mathbf{\Sigma}_3^R$	0.18	0.87	0.91	0.12	0.98	0.94	0.08	1.04	0.97
$\mathbf{\Sigma}_3^*$	$\mathbf{\Sigma}_2^H$	$\mathbf{\Sigma}_3^R$	0.21	1.01	0.95	0.13	1.07	0.97	0.09	1.12	0.97
$\mathbf{\Sigma}_3^{**}$	$\mathbf{\Sigma}_2^H$	$\mathbf{\Sigma}_3^R$	0.21	1.01	0.95	0.13	1.09	0.97	0.09	1.14	0.97
$\mathbf{\Sigma}_3^*$	$\mathbf{\Sigma}_2^R$	$\mathbf{\Sigma}_3^R$	0.21	1.05	0.95	0.13	1.08	0.97	0.09	1.12	0.97
$\mathbf{\Sigma}_3^{**}$	$\mathbf{\Sigma}_2^R$	$\mathbf{\Sigma}_3^R$	0.22	1.06	0.96	0.14	1.10	0.97	0.09	1.15	0.97
					Proportional						
$\mathbf{\Sigma}_3$		$\mathbf{\Sigma}_3^H$	0.19	1.04	0.97	0.14	1.16	0.98	0.10	1.25	0.99
$\mathbf{\Sigma}_3^*$	$\mathbf{\Sigma}_2^H$	$\mathbf{\Sigma}_3^H$	0.21	1.14	0.98	0.15	1.24	0.98	0.10	1.33	0.99
$\mathbf{\Sigma}_3^{**}$	$\mathbf{\Sigma}_2^H$	$\mathbf{\Sigma}_3^H$	0.21	1.11	0.97	0.14	1.22	0.98	0.10	1.31	0.99
$\mathbf{\Sigma}_3$		$\mathbf{\Sigma}_3^R$	0.17	0.92	0.94	0.12	1.00	0.95	0.08	1.07	0.96
$\mathbf{\Sigma}_3^*$	$\mathbf{\Sigma}_2^H$	$\mathbf{\Sigma}_3^R$	0.19	1.04	0.95	0.13	1.10	0.97	0.09	1.17	0.98
$\mathbf{\Sigma}_3^{**}$	$\mathbf{\Sigma}_2^H$	$\mathbf{\Sigma}_3^R$	0.19	1.00	0.95	0.13	1.07	0.97	0.09	1.14	0.98
$\mathbf{\Sigma}_3^*$	$\mathbf{\Sigma}_2^R$	$\mathbf{\Sigma}_3^R$	0.20	1.07	0.95	0.13	1.11	0.98	0.09	1.17	0.98
$\mathbf{\Sigma}_3^{**}$	$\mathbf{\Sigma}_2^R$	$\mathbf{\Sigma}_3^R$	0.20	1.04	0.95	0.13	1.08	0.97	0.09	1.15	0.98

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

Table A2: Comparison of the different variance estimators for high separation,for the 3 sample sizes separatly for one parameter, β_{13} for modal and proportional assignment

Final	Components		500			1000			2000		
	2^{nd} step	3^{rd} step	se	se/sd	coverage	se	se/sd	coverage	se	se/sd	coverage
					Modal						
$\mathbf{\Sigma}_3$		$\mathbf{\Sigma}_3^H$	0.14	0.93	0.95	0.10	1.08	0.97	0.07	1.09	0.96
$\mathbf{\Sigma}_3^*$	$\mathbf{\Sigma}_2^H$	$\mathbf{\Sigma}_3^H$	0.14	0.94	0.95	0.10	1.09	0.97	0.07	1.10	0.96
$\mathbf{\Sigma}_3^{**}$	$\mathbf{\Sigma}_2^H$	$\mathbf{\Sigma}_3^H$	0.15	0.95	0.95	0.10	1.10	0.98	0.07	1.11	0.96
$\mathbf{\Sigma}_3$		$\mathbf{\Sigma}_3^R$	0.14	0.94	0.94	0.10	1.10	0.97	0.07	1.09	0.96
$\mathbf{\Sigma}_3^*$	$\mathbf{\Sigma}_2^H$	$\mathbf{\Sigma}_3^R$	0.15	0.96	0.95	0.10	1.10	0.98	0.07	1.11	0.96
$\mathbf{\Sigma}_3^{**}$	$\mathbf{\Sigma}_2^H$	$\mathbf{\Sigma}_3^R$	0.15	0.96	0.95	0.10	1.10	0.96	0.07	1.11	0.96
$\mathbf{\Sigma}_3^*$	$\mathbf{\Sigma}_2^R$	$\mathbf{\Sigma}_3^R$	0.15	0.96	0.95	0.10	1.10	0.97	0.07	1.11	0.96
$\mathbf{\Sigma}_3^{**}$	$\mathbf{\Sigma}_2^R$	$\mathbf{\Sigma}_3^R$	0.15	0.96	0.95	0.10	1.10	0.98	0.07	1.11	0.96
					Proportional						
$\mathbf{\Sigma}_3$		$\mathbf{\Sigma}_3^H$	0.15	0.95	0.95	0.10	1.14	0.98	0.07	1.11	0.97
$\mathbf{\Sigma}_3^*$	$\mathbf{\Sigma}_2^H$	$\mathbf{\Sigma}_3^H$	0.15	0.96	0.95	0.10	1.15	0.98	0.07	1.15	0.97
$\mathbf{\Sigma}_3^{**}$	$\mathbf{\Sigma}_2^H$	$\mathbf{\Sigma}_3^H$	0.15	0.95	0.95	0.10	1.14	0.98	0.07	1.14	0.97
$\mathbf{\Sigma}_3$		$\mathbf{\Sigma}_3^R$	0.14	0.92	0.95	0.10	1.10	0.96	0.07	1.10	0.96
$\mathbf{\Sigma}_3^*$	$\mathbf{\Sigma}_2^H$	$\mathbf{\Sigma}_3^R$	0.14	0.93	0.95	0.10	1.10	0.96	0.07	1.11	0.96
$\mathbf{\Sigma}_3^{**}$	$\mathbf{\Sigma}_2^H$	$\mathbf{\Sigma}_3^R$	0.14	0.92	0.95	0.10	1.10	0.96	0.07	1.10	0.96
$\mathbf{\Sigma}_3^*$	$\mathbf{\Sigma}_2^R$	$\mathbf{\Sigma}_3^R$	0.14	0.93	0.95	0.10	1.10	0.96	0.07	1.11	0.96
$\mathbf{\Sigma}_3^{**}$	$\mathbf{\Sigma}_2^R$	$\mathbf{\Sigma}_3^R$	0.14	0.92	0.95	0.10	1.10	0.96	0.07	1.10	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