DETERMINING THE NUMBER OF CLASSES IN LATENT CLASS REGRESSION MODELS

DETERMINING THE NUMBER OF CLASSES IN LATENT CLASS REGRESSION MODELS

BY

SHERRY LUO, B.Sc.

A THESIS
SUBMITTED TO THE DEPARTMENT OF MATHEMATICS & STATISTICS
AND THE SCHOOL OF GRADUATE STUDIES
OF MCMASTER UNIVERSITY
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

Master of Science (2021)	McMaster University
(Mathematics & Statistics)	Hamilton, Ontario, Canada

TITLE:	Determining the number of classes in latent class regression models
AUTHOR:	Sherry Luo B.Sc. (Statistics), McMaster University, Hamilton, Canada
SUPERVISOR:	Dr. Paul D. McNicholas
NUMBER OF PAGES:	xi, 92

Abstract

Latent class regression (LCR) is a statistical method used to identify qualitatively different groups or latent classes within a heterogeneous population and commonly used in the behavioural, health, and social sciences. Despite the vast applications, an agreed fit index to correctly determine the number of latent classes is hotly debated. To add, there are also conflicting views on whether covariates should or should not be included into the class enumeration process. We conduct a simulation study to determine the impact of covariates on the class enumeration accuracy as well as study the performance of several commonly used fit indices under different population models and modelling conditions. Our results indicate that of the eight fit indices considered, the aBIC and BLRT proved to be the best performing fit indices for class enumeration. Furthermore, we found that covariates should not be included into the enumeration procedure. Our results illustrate that an unconditional LCA model can enumerate equivalently as well as a conditional LCA model with its true covariate specification. Even with the presence of large covariate effects in the population, the unconditional model is capable of enumerating with high accuracy. As noted by Nylund-Gibson and Masyn (2016), a misspecified covariate specification can easily lead to an overestimation of latent classes.

Therefore, we recommend to perform class enumeration without covariates and determine a set of candidate latent class models with the aBIC. Once that is determined, the BLRT can be utilized on the set of candidate models and confirm whether results obtained by the BLRT match the results of the aBIC. By separating the enumeration procedure of the BLRT, it still allows one to use the BLRT but reduce the heavy computational burden that is associated with this fit index. Subsequent analysis can then be pursued accordingly after the number of latent classes is determined.

Acknowledgements

First and foremost I am deeply grateful to my supervisor, Paul McMicholas. Without your support, I would not have had the confidence to enter into this program and try. Your patience and understanding has been immensely helpful. It has allowed me to reflect and act on what is important to me and my future. You have been a great source of professional support by always letting me explore my own ideas and willing to advise and help when necessary. I am in a better position today than I was when I first met you, because you said yes to being my supervisor. Thank you Paul.

I would also like to thank Dr. Pratheepa Jeganathan and Dr. Nicholas Bock for examining my thesis and the helpful suggestions.

Mom and dad, you have provided me everything I need to function successfully in this crazy world. All my past and future experiences is due to the sacrifice you both made over 25 years ago. Michelle, Steven and I are very grateful. Michelle and Steven, my favourite siblings, thanks for always putting up with my bull **** ha.

To my mentor, of the bajillion people out there, you inspire me the most. Whenever I had doubts or felt beaten up because of x,y,z you would always have some words of wisdom or story to lift me up. The opportunity to have worked and learned from you were definitely my highlights of 2020.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Latent Variable Models

Latent variable models (LVM) are popular statistical models commonly used in the behavioural, health and social sciences. In such fields, hypothetical constructs such as intelligence, social status, and happiness are regarded as “hidden” or “latent”, since these constructs are not directly observable or measurable in the population (Everitt, 1984). Rather it is assumed that the latent variables can be indirectly measured by a set of observable variables known as manifest variables or items.

Table 1.1 is a unified framework of LVM presented by Muthén (2008). LVM can be classified according to the type of analysis and whether the latent variables are of strictly categorical, strictly continuous, or of mixed type. Each entry of Table 1.1 is a distinct LVM model, with rows separating models into cross-sectional and longitudinal models and columns corresponding to the metric level of the latent variables. Those interested in an in-depth overview of the models presented in Table 1.1 can refer to Muthén (2008).

Table 1.1: Classification of latent variable models as presented by Muthén (2008).

	Latent Variables		
Model Type	Continuous	Categorical	Hybrids
Cross Sectional	Factor analysis, SEM	Regression mixture analysis, Latent Class Analysis	Factor mixture analysis
Longitudinal	Growth analysis	Latent transition analysis, Latent class growth analysis	Growth mixture models

In this thesis, we study categorical latent variables of cross sectional nature, specifically latent class analysis (LCA) models. The objective of LCA is to classify respondents into homogeneous subgroups or latent classes based on their observed responses to a set of items that may be categorical or continuous. Generally, latent profile analysis is used when items are continuous and latent class analysis when items are categorical. Applications of LCA can be found in various substantive research areas such as health, behavioural, and social sciences. Some interesting examples include relating eating disorders patients to mortality rates (Crow *et al.*, 2012) and patterns of acculturation among Asian Americans (Jang *et al.*, 2017).

LCA was first brought to light in the social and behavioural sciences by Lazarsfeld and Henry (1968). Though they provided a comprehensive and detailed mathematical treatment of this topic, it remained in the shadow for almost a decade because the parameter estimates were difficult to obtain at the time. This changed when Goodman (1974) developed a straightforward and readily implementable method for obtaining maximum likelihood of latent class model parameters. Presently, the expectation-maximization (EM) algorithm (Dempster *et al.*, 1977) is commonly used to compute the parameter estimates and readily available in popular statistical softwares such as `Latent Gold` (Vermunt and Magidson, 2016) and `Mplus` (Muthén and Muthén, 2017).

Many extensions of LCA have been accomplished in the last few decades. The latent class model became more flexible when it was introduced in a log-linear modelling framework (Haberman, 1974; Hagenaars, 1998). This paved the way for several new developments such as multilevel LCA (Vermunt, 2003, 2008), LCA for longitudinal data (Chung *et al.*, 2014), Bayesian LCA (White and Murphy, 2014), and most prominently, inclusion of covariates in LCA (Dayton and Macready, 2012; Huang and Bandeen-Roche, 2004) or Latent Class Regression (LCR). LCR is increasingly being used as an analytic tool since it not only allows the researcher to build a robust classification model but the inclusion of covariates can improve the prediction of class membership and aid the identification of the latent classes (Dayton and Macready, 2012; Hagenaars, 1993).

Despite the vast literature and growing applications related to LCA, determining the correct number of classes, also known as class enumeration, remains an unresolved issue. Even under the assumption that the population is indeed heterogeneous *a priori*, hypotheses regarding the exact number or nature of the sub-populations are rarely known (Masyn, 2013). In general, class enumeration is typically done with combination of substantive theory and examining a set of fit indices. This is an extremely laborious task as it requires considering a set of models with varying number of classes then observing and comparing each individual fit index to determine the optimal number of classes. Several simulation studies have examined the performance of different fit index under different modelling conditions (Nylund *et al.*, 2007; Peugh and Fan, 2015; Morovati, 2014). However, majority of these studies have delivered mixed results, thus, it is difficult to conclude on a preferable fit index for enumerating the correct number of latent classes.

To add more nuance, there are conflicting views on when to include covariates during the class enumeration process (Nylund-Gibson and Masyn, 2016). In latent variable modelling literature, there are two contrasting views. Some suggest that it's best to first perform enumeration without covariates. More specifically, determine the number of latent classes using an unconditional LCA model (Collins and Lanza, 2010; Tofighi *et al.*, 2008; Masyn, 2013). Others advocate to include covariates simultaneously during the enumeration process as the additional information may increase the performance of the selected fit index (Lubke and Muthén, 2007; Li and Hser, 2011; Peugh and Fan, 2015). These conflicting views have led to inconsistent practices among researchers applying LCA (Nylund-Gibson and Masyn, 2016). Thus we conduct a simulation study examining how the performance of fit indices are affected with presence of covariate effects and provide more insight on whether covariates aid or hinder in determining the number of latent classes.

The paper will be organized in the following manner: Chapter 2 is a literature review and introduces the LCA model and class enumeration. Additionally we provide theory behind the fit indices considered in the study. The full simulation details can be found in Chapter 3. Results of the simulation study are explained in Chapter 4 and final remarks and conclusions in Chapter 5.

Chapter 2

Literature Review

2.1 The Latent Class Model

Suppose there are M categorical response items and each item u_m, where $m = 1, 2 \ldots, M$, has $r_{u_m} = 1, 2, \ldots, R_{u_m}$ response categories. Let the response of subject i on item u_m be denoted by $u_{mi} = r_{u_{mi}}$, and the full response vector by $\boldsymbol{u_i} = [u_{1i}, \ldots, u_{Mi}]^\intercal$. The unconditional LCA model is represented in Figure 2.1 (Lazarsfeld and Henry, 1968; Collins and Lanza, 2010). It assumes the M items are reflective of an underlying categorical latent variable c with K latent classes such that $c = k; k = 1, 2, ..., K$. The LCA model of $P(\boldsymbol{u_i})$ is expressed as

$$P(\boldsymbol{u_i}) = \sum_{k=1}^{K} \Big[P(c_i = k) P(u_{1i}, u_{2i}, \ldots, u_{Mi} | c_i = k) \Big]. \tag{2.1.1}$$

Equation (2.1.1) can be simplified by *local independence* which is typically assumed on the M items conditional on latent class membership. This assumption implies that all the associations shared among the observed items is strictly explained by the latent variable or another way of putting it, the latent variable explains why the observed items are related to each other (Nylund-Gibson and Choi, 2018). Under the local independence assumption, we have

$$
\begin{aligned}
P(\boldsymbol{u_i}) &= \sum_{k=1}^{K} P(c_i = k)P(u_{1i}|c_i = k)P(u_{2i}|c_i = k) \times \ldots \times P(u_{Mi}|c_i = k) \\
&= \sum_{k=1}^{K} P(c_i = k) \times \left[\prod_{m=1}^{M} \prod_{r_{u_m}=1}^{R_{u_m}} P(u_{mi} = r_{u_m}|c_i = k)^{1(u_{mi}=r_{u_m})} \right],
\end{aligned}
\tag{2.1.2}
$$

where $1(u_{mi} = r_{u_m}) = 1$ if $u_{mi} = r_{u_m}$ and 0 otherwise. Local independence is depicted in Figure 2.1. As shown, the items, denoted in squares, are only connected by the latent variable, denoted in the circle, as indicated by the one-way directional path which signifies that the M items are only related through the latent variable.

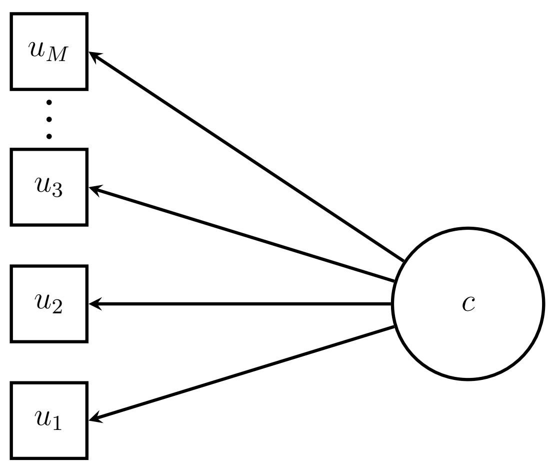

Figure 2.1: A classic representation of the unconditional latent class model. The latent variable c, enclosed in the circle, is measured by a set of items $u_1, u_2, \ldots, u_m$, enclosed in the rectangles. One-direction arrows connect the items and latent variable but no arrows connect the items, illustrating conditional independence on the items.

There are two main sets of parameters that describe the LCA model. First, the relative size of each latent class is described by the class proportion probabilities $P(c_i = k)$. Class proportion probabilities are the structural parameters of the LCA model and defines the distribution of the latent variable. Latent classes are assumed to be mutually exclusive and exhaustive such that individuals may belong to one and only one of the K latent class, thus, $\sum_{k=1}^{K} P(c_i = k) = 1$. The distribution of the class membership probabilities is expressed as intercepts of a multinomial logistic model

$$P(c_i = k) = \frac{\exp(\gamma_{0k})}{1 + \sum_{h=1}^{K-1} \exp(\gamma_{0h})}, \tag{2.1.3}$$

where $\gamma_{0K} = 0$ and K is the reference category and represents the odds of membership in latent class k relative to reference latent class K. The choice of the reference category is arbitrary and will not affect the results, however, it can impact the ease of interpretation.

Now, the relationship between the observed items and latent variable is captured by the set of item-response probabilities $P(u_{mi} = r_{u_m} | c_i = k)$. They are the measurement parameters of model and measure how likely an individual will endorse a particular item u_m with response r_{u_m} within each latent class. Much like factor analysis, the overall distribution of the item-response probabilities help investigators assign meaning to the latent classes during the interpretation phase of the analysis. In practice, binary and ordinal items are commonly used in LCA applications (Nylund-Gibson and Choi, 2018). Different parameterizations can be used to express the item-response probabilities. The most widely used parameterizations that result in equivalent models include the probability parameterization, the log-linear parameterization, and the logistic parameterization Masyn (2017). As we are using `Mplus`, a latent variable modelling program with a wide variety of analysis capabilities (Muthén and Muthén, 2017), the item-response probabilities are parameterized under a logistic regression framework, specifically a latent response variable formulation.

Assume that each item-response u_{mi} arises from a continuous latent response variable $u_{mi}*$, or stated differently, $u_{mi}*$ is an underlying continuum of u_{mi} (Agresti, 2002; Masyn *et al.*, 2014) and the relationship can be modelled by

$$u^*_{mi} = \mu_{mi|\boldsymbol{x_i}} + \epsilon_i, \tag{2.1.4}$$

where $\mu_{mi|\boldsymbol{x}} \equiv E[u^*_{mi}|\boldsymbol{x_i}] = \boldsymbol{x_i^\intercal \beta_{u_m}}$ such that $\boldsymbol{x_i} = [x_{1i}, \ldots, x_{pi}]^\intercal$ and $\boldsymbol{\beta_{u_m}} = [\beta_1, \ldots, \beta_p]$ are the related regression parameters. Let ϵ_i be the error term that follows a standard logistic distribution. Additionally suppose we have cut-off points or thresholds on the $u_{mi}*$ scale such that $\tau_{0k} < \tau_{1k} < \tau_{1k} < \tau_{2k} < \ldots < \tau_{R_{u_m}-1k} < \tau_{R_{u_m}k}$ where $\tau_{0k} = -\infty$ and $\tau_{R_{u_m}k} = \infty$. The relationship between $u_{mi}*$ and u_{mi} is given by

$$u_{mi} = \begin{cases} 1 & \text{if } -\infty < u^*_{mi} \leq \tau_{1k}, \\ 2 & \text{if } \tau_{1k} < u^*_{mi} \leq \tau_{2k}, \\ \vdots & \qquad \vdots \\ R_{u_m} - 1 & \text{if } \tau_{R_{u_m}-2k} < u^*_{mi} \leq \tau_{R_{u_m}-1k}, \\ R_{u_m} & \text{if } \tau_{R_{u_m}-1k} < u^*_{mi} \leq \infty \end{cases} \tag{2.1.5}$$

where each interval corresponds to a range that a response category may fall into. Note that for the unconditional LCA model, (2.1.4) is $u^*_{mi} = \epsilon_i$ since $\mu_{mi|\boldsymbol{x_i}} = 0$.

Thus the item-response probabilities are given by

$$\begin{aligned} P(u_{mi} = r_{u_m}|c_i = k) &= P(\tau_{r_{u_m}-1k} < u^*_{mi} \leq \tau_{r_{u_m}k}) \\ &= P(u^*_{mi} \leq \tau_{r_{u_m}k}) - P(u^*_{mi} \leq \tau_{r_{u_m}-1k}) \\ &= F_\epsilon(\tau_{r_{u_m}k}) - F_\epsilon(\tau_{r_{u_m}-1k}), \end{aligned} \tag{2.1.6}$$

where $P(u_{mi} \leq r_{u_m}|c_i = k) = F_\epsilon(\tau_{r_{u_m}k})$, is the cumulative distribution of the standard logistic distribution. The item-response probabilities are expressed as the difference in cumulative probabilities. With ordinal items, we are interested in obtaining the probability of being at or above a response category

$$\begin{aligned} P(u_{mi} \geq r_{u_m}|c_i = k) &= 1 - P(u_{mi} < r_{u_m}|c_i = k) \\ &= 1 - P(u^*_{mi} < \tau_{r_m k}) \\ &= 1 - F_\epsilon(\tau_{r_{u_m}k}) \\ &= \frac{1}{1 + \exp(\tau_{r_{u_m}k})}. \end{aligned} \tag{2.1.7}$$

Applying a logit transformation, we obtain

$$\begin{aligned} \text{logit}[P(u_{mi} \geq r_{u_{r_m}}|c_i = k)] &= \log\left[\frac{P(u_{mi} \geq r_{u_m}|c_i = k)}{P(u_{mi} < r_{u_m}|c_i = k)}\right] \\ &= -\tau_{r_{u_m}k}, \end{aligned} \tag{2.1.8}$$

where $\tau_{r_{u_m}k}$ is equal to the negative log-odds of a response on item u_m that is greater than or equal to r_{u_m} versus responding to lower categories within latent class k i.e., the odds of being at or above a specific response category is higher when the threshold is more negative. We see that the latent response formulation is equivalent to a binary logistic regression in which response categories $1, 2, \ldots, r_{u_m}$ form a single category and the remaining response categories $(r_{u_m} + 1), \ldots, R_{u_m}$ form another category (Agresti, 2002).

2.2 Latent Class Regression Model

Thus far we have discussed the unconditional LCA model where the relationships between the observed items is solely explained by the latent variable. In practice, however, there may be applications where we would like to relate a set of p covariates, $\boldsymbol{x} = [x_1, x_2, \dots, x_p]^\intercal$, to the latent variable and the observed items. This may improve the prediction of class membership and facilitate in the identification and interpretation of latent classes (Park and Yu, 2018). In the following we will extend the unconditional LCA model to incorporate covariates.

2.2.1 Covariate Pathways in LCR

Covariates can influence a set of observed items indirectly via the latent variable, or directly in which the latent variable is entirely avoided (Masyn, 2013). Figure 2.2 illustrate examples of LCR path diagrams with a single latent variable c that are commonly seen in application. An indirect pathway is depicted in Figure 2.2a. In this case, the class membership probabilities are conditional on a set of covariates resulting the item-response probabilities to be indirectly influenced through the latent variable. Extending (2.1.2), the conditional latent class model with indirect effects is given by

$$P(\boldsymbol{u_i}|\boldsymbol{x_i}) = \sum_{k=1}^{K} P(c_i = k|\boldsymbol{x_i}) \times \left[\prod_{m=1}^{M} \prod_{r_{u_m}=1}^{R_{u_m}} P(u_{mi} = r_{u_m}|c_i = k)^{1(u_{mi}=r_{u_m})} \right], \quad (2.2.1)$$

where $1(u_{mi} = r_{u_m}) = 1$ if $u_{mi} = r_{u_m}$ and 0 otherwise. The item-response probabilities remain unchanged since they are independent of the covariates conditional on latent class membership. The class membership probabilities are parameterized using a multinomial regression such that

$$P(c_i = k|\boldsymbol{x}_i) = \frac{\exp(\gamma_{0k} + \sum_{j=1}^{p} \gamma_{jk} x_{ij})}{1 + \sum_{h=1}^{k-1} \exp(\gamma_{0k} + \sum_{j=1}^{p} \gamma_{jk} x_{ij})}, \tag{2.2.2}$$

where $\gamma_{0K} = \gamma_{jK} = 0$, $j = 1, 2, .., p$ for identification. Equivalently,

$$\begin{aligned} \text{logit}\big[P(c_i = k|\boldsymbol{x}_i)\big] &= \log\left[\frac{P(c_i = k|\boldsymbol{x_i})}{P(c_i = K|\boldsymbol{x_i}}\right] \\ &= \gamma_{0k} + \sum_{j=1}^{p} \gamma_{jk} x_{ij}. \end{aligned} \tag{2.2.3}$$

Exponentiating both sides of (2.2.3) we can interpret the odd as a function of the covariates. More precisely, exponentiating the intercept γ_{0k} represent the odds of membership in latent class k relative to reference latent class K when covariates are zero. Exponentiating the regression parameters γ_{jk} can be interpreted as the change in the odds of membership in latent class k in relation to the reference class K associated with a one unit change in x_{ij} while fixing the other covariates constant.

Figure 2.2b models a direct pathway where covariates influence item u_1 directly and bypass the latent variable. With both indirect and direct pathways as shown in Figure 2.2c, the latent variable depends on the covariates, and with the exception of item u_1, all other items are conditionally independent of the covariates given latent class membership. The direct pathway on u_1 implies measurement non-invariance, that is, individuals in the same class differ in their response probability for item u_1 as a function of x-values. In other words, the conditional model with an indirect and

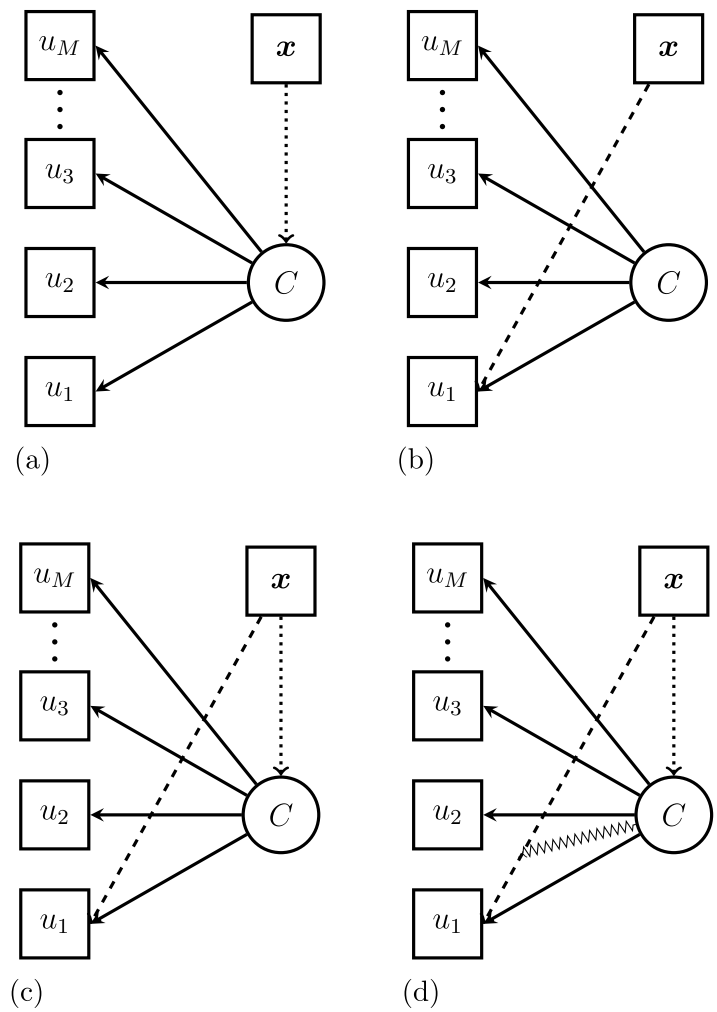

Figure 2.2: The various pathways seen in latent class regression models (also known as the conditional LCA model) where panel (A) represents an indirect effect on the items through latent variable,(B) direct effect between the covariates $\boldsymbol{x}$ and u_1 where direct effects are an indication of measurement non-invariance, (C) $\boldsymbol{x}$ has an indirect effect on c and a direct effect on u_1 (D) effects similar to panel C, but, the direct effect on u_1 is class varying

direct effect is

$$P(\boldsymbol{u_i}|\boldsymbol{x_i}) = \sum_{k=1}^{K} \Big[P(c_i = k|\boldsymbol{x_i}) \times \prod_{r_{u_1}=1}^{R_{u_1}} P(u_{1i} = r_{u_1}|c_i = k, \boldsymbol{x}_i)^{1(u_{1i}=r_{u_1})} \times \prod_{m=2}^{M} \prod_{r_{u_m}=1}^{R_{u_m}} P(u_{mi} = r_{u_m}|c_i = k)^{1(u_{mi}=r_{u_m})} \Big]. \tag{2.2.4}$$

The set of covariates $\boldsymbol{x}$, is related u_1 via the latent response formulation in (2.1.4) where $\mu_{1i|\boldsymbol{x_i}} = \boldsymbol{x}^\intercal \boldsymbol{\beta_{u_1}}$. Thus, the item-response probability of u_{1i} conditional on latent class k is given by

$$P(u_{1i} \geq r_{u_1}|c_i = k, \boldsymbol{x_i}) = \frac{1}{1 + exp(\tau_{r_{u_1}k} - \boldsymbol{x_i}^\intercal \boldsymbol{\beta}_{u_1})} \tag{2.2.5}$$

or, equivalently,

$$\log\left(\frac{P(u_{1i} \geq r_{u_1}|c_i = k, \boldsymbol{x_i}}{P(u_{1i} < r_{u_1}|c_i = k, \boldsymbol{x_i})} \right) = -\tau_{r_{u_1}k} + \boldsymbol{x_i}^\intercal \boldsymbol{\beta}_{u_1}. \tag{2.2.6}$$

From (2.2.6), each response category has its own threshold parameter $\tau_{r_{u_1}k}$ but the effect of $\boldsymbol{\beta_{u_1}} = [\beta_1, \ldots, \beta_p]$ is identical and fixed for across each category and not class-specific for $c = 1, 2, \ldots, K$. That is, the same proportionality constant is applied for different response categories such that the direct effect of x_j will be the same for every latent class k. It is not class-varying. This is known as the proportional odds assumption (McCullagh, 2005) because for a fixed $\boldsymbol{x} = \boldsymbol{x_1}$ and $\boldsymbol{x} = \boldsymbol{x_2}$, it satisfies

$$\text{logit}\big[P(u_{1i} \geq r_1|c_i = k, \boldsymbol{x_2})\big] - \text{logit}\big[(P(u_{1i} \geq r_1|c_i = k, \boldsymbol{x_1})\big] = \boldsymbol{\beta}_{u_1}(\boldsymbol{x_2} - \boldsymbol{x_1})^\intercal \ \forall \ c{=}1,2,\ldots,K, \tag{2.2.7}$$

where $\boldsymbol{\beta_{u_1}}$ is the proportionality constant. Thus the odds of responding at or above category, r_{u_1}, at $\boldsymbol{x} = \boldsymbol{x_1}$ is $\exp\left[\boldsymbol{\beta}_1(\boldsymbol{x_2} - \boldsymbol{x_1})^\intercal\right]$ times the odds at $\boldsymbol{x} = \boldsymbol{x_2}$ given membership in latent class k for all $k = 1, 2, \ldots K$.

In contrast Figure 2.2d incorporates an indirect and class-varying direct effect which is indicated by the zig-zag arrow between the latent variable and direct effect path from $\boldsymbol{x}$ to u_1. With class-varying, a k subscript is added to the proportionality constant in (2.2.6) such that

$$\log\left(\frac{P(u_{1i} \geq r_1 | c_i = k, \boldsymbol{x_i})}{P(u_{1i} < r_1 | c_i = k, \boldsymbol{x_i})}\right) = -\tau_{r_{u_1}k} + \boldsymbol{x_i}^\intercal \boldsymbol{\beta}_{u_1 k} \quad \text{for} \quad c\mathit{=1,2,\ldots,K}. \tag{2.2.8}$$

Each logit has a different threshold $\tau_{r_{u_1}k}$ but the direct effect of $\boldsymbol{x_i}$ on u_1 is class-specific where $\boldsymbol{\beta}_{u_1 k} = [\beta_{1k}, \ldots, \beta_{pk}]$.

2.2.2 Consequences of ignoring direct effects

When comparing two or more classes it is important to determine whether each item measures the latent variable in the same manner for different classes. This assumption is called measurement invariance. A violation of this assumption implies that a specific item is not measuring the same characteristic across each class. In LCA models, a covariate x is a source of measurement non-invariance on an item u_m if a direct effect is present between x and u_m (Masyn, 2017). More specifically, the probability of endorsing item u_m for all individuals belonging in latent class k would differ based on their individual x value. But the probabilities of endorsing $u_1, u_2, \ldots u_{m-1}, u_{m+1}, \ldots, u_M$ would be equal for all individuals belonging in latent class k since these items are conditionally independent of x. A long standing approach of identifying direct effects is to regress each item on the covariates and conduct separate LCR analysis.

Each conditional model is compared to the unconditional model (baseline model with no covariates) using a significance test. In other words, DEs are determined to be present when the direct pathway between covariates and item in question is statistically significant. Other approaches for larger, complex LCA models are summarized in Janssen *et al.* (2019). If direct effects are ignored between u_m and x, the resultant estimates of the measurement model may be biased because of the unmodelled residual association between the item and covariate (Janssen *et al.*, 2019). As a result, this will lead to misspecfication of the latent classes. To illustrate this point, Masyn (2017) examined well-separated and highly homogeneous population models with direct effect sizes. The simulation study concluded that even minimal direct effects could lead to substantial bias in the model parameters. Janssen *et al.* (2019) examined the stability and bias in parameter estimates when direct effects were excluded from the analysis. The study concluded that direct effects were necessary when the relationship between the covariates and items were strongly associated or when the measurement model was overall weak. When direct effects were not modelled in these situations, the bias of the estimates increased and coverage decreased severely. Therefore, covariates that contribute to direct effects or indirect effects cannot be ignored from the modelling procedure and must always be incorporated to ensure model estimates are unbiased. Thus, it is important to determine whether covariates are incorporated with the enumeration process or after the number of classes has been decided. In the next section details of the most commonly used fit indices in latent variable modelling are discussed.

2.3 Selecting the best model

Selecting the correct number of latent classes, also known as class enumeration, is one of the major challenges in latent variable modelling. Generally, the enumeration phase is a time-consuming process as it requires estimation of several competing models with varying number of classes, and difficult since the final model is selected based on the examination of several fit criteria (Masyn, 2013). Several simulations studies have examined this issue under various types of latent variable models and modelling conditions. Additionally it is further complicated by the number of possible fit indices, which vary in sensitivity to sample size, model complexity and the presence of covariates (Hu *et al.*, 2017). Despite the numerous suggestions offered in literature, a unanimous and preferable fit index for deciding the number of latent classes remains an unresolved issue. Generally these model selection techniques can be grouped into three categories: likelihood ratio based tests, information criteria, and classification based criterion.

2.3.1 Likelihood Ratio Test

The likelihood ratio test (LRT) determines whether an additional latent class significantly improves model fit. More specifically we test the null hypothesis

$$H_0\colon \text{ Number of classes} = k - 1 \tag{2.3.1}$$

versus the alterantive hypothesis

$$H_1\colon \text{ Number of classes} = k. \tag{2.3.2}$$

The LRT statistic is defined as the difference in the log likelihoods:

$$LRT = -2(LL_{k-1} - LL_k), \tag{2.3.3}$$

where LL_{k-1} and LL_k are the maximum likelihood estimates of the null and alternative models, respectively. Most often with nested models, under certain regularity conditions, the LRT statistic assumes a chi-square distribution with degrees of freedom equal to the difference in the number of parameters of the two models. Unfortunately the LRT cannot be used in the same manner in LCA models because the usual regularity conditions associated with the chi-square LRT do not apply for comparing LCA models with different number of classes (Tekle *et al.*, 2016). A $(k-1)$-class model can be obtained by restricting one of the class membership probabilities of a k-class model to zero. As a consequence, we are fixing parameters at the boundary of its permissible parameter space rather than in its impermissible parameter space. Everitt (1988) and Nylund *et al.* (2007) conducted simulations examining the performance of the chi-square LRT and both concluded inflated Type I error rates in their studies.

To overcome this issue, various modified LRTs have been proposed for latent variable models. One such example is the parametric bootstrapped likelihood ratio test (BLRT; McLachlan and Peel (2000)). The BLRT uses the same test statistic as chi-square LRT, however, BLRT obtains a p-value by using bootstrap samples to estimate the sampling distribution of the LRT. The BLRT is calculated as follows (Asparouhov and Muthén, 2012):

1. In the k-class run, compute LRT (2.3.3) by estimating the parameters of the k-class model and $(k-1)$-class model.

2. Generate a bootstrap sample using the paramters under null hypothesis model as the true population values. Analyze the bootstrap asmple under the k-class model and estimate the new LRT.

3. Repeat step 2 several times to construct the sampling distribution of the LRT.

4. Estimate the p-value by comparing the distribution obtained in step 3 with the LRT obtained in step 1.

The p-value is the proportion of bootstrap LRT values that is larger or equal to the LRT value from step 1. It is used to decide whether the (k-1)-class model should be rejected in favour of the k-class model. It is apparent that the BLRT is computationally intensive as it requires fitting an LCA model multiple times to each bootstrap sample, which can take a few seconds to hours to compute for more complex models (Dziak *et al.*, 2014). Despite this, BLRT is worth studying since it is widely used class enumeration tool for latent variable models and shown promising results in some studies.

As an alternative to BLRT, Lo (2001) proposed the Lo-Mendel-Rubin LRT (LMR). The LMR-LRT extends the work of Vuong (1989) and provides an approximation of the LRT distribution to allow comparisons between Gaussian mixture models that differ in the number of classes. Let $f(u;\theta)$ and $g(u;\lambda)$ represent the density of a k-class model and $(k-1)$-class model respectively, where $\boldsymbol{u_i}$ is a vector of observed items and θ and λ are the model parameters. The LMR is

$$\text{LMR} = LL_f(\hat{\theta}) - LL_g(\hat{\lambda}) = \sum_{i=1}^{n} \log \frac{f(\boldsymbol{u_i};\hat{\theta})}{g(\boldsymbol{u_i};\hat{\lambda})}. \tag{2.3.4}$$

The distribution is discussed in Lo (2001). However, the LMR exhibited high Type

1 errors during their simulation study and proposed an adjusted LMR to account for the high Type I error rates. The adjusted LMR is

$$\text{aVLMR} = \frac{\text{LMR}}{1 + [(p-q)\log(n)]^{-1}}, \tag{2.3.5}$$

where p and q are equal to the number of estimated parameters for the models with degrees of freedom equal to $(p-q)$.

Jeffries (2003) pointed a mathematical flaw in Lo (2001) paper suggesting that one of the conditions proposed for aVLMR is incorrect. Regardless, the aVLMR has still been widely used and shown to be effective in recovering the number of latent classes. Both the BLRT and aVLMR test whether the $(k-1)$-class should be rejected in favour of the k-class model. More specifically, a significant aVLMR and BLRT (e.g. $p \leq 0.05$) implies that the $(k-1)$-class model is rejected in favour of the k-class model. In contrast a non-significant aVLMR and BLRT (e.g. $p > 0.05$) implies that $(k-1)$-class model fits the data as well as k-class model, thus supporting the $(k-1)$-class model.

2.3.2 Information Criteria

Perhaps the most popular fit indices are information criteria (IC). Unlike LRT, they allow for comparison of multiple models that may or may not be nested. Information criteria were originally proposed by Akaike (1973) who utilized the Kullback-Leibler (KL) Divergence measure to form the Akaike IC (AIC) which tries to select the model that minimizes the KL Divergence. Later on Schwarz (1978) incorporated Bayesian statistics to form the Bayesian IC (BIC) that attempts to select the model with the highest posterior probability. Since then, many information criteria have been proposed based on different or similar theoretical frameworks as the AIC and BIC. An information criterion can be summarized in terms of its log-likelihood function and an added penalty term which discourages over fitting of the data. It is expressed as

$$\text{IC} = -2LL + \text{penalty}, \tag{2.3.6}$$

where LL is the maximized log likelihood value for the model in consideration. The *penalty* term measures the complexity of the model by taking account sample size and the number of parameters being estimated in the model.

Although IC are motivated by different frameworks and goals, algebraically, the difference in penalty terms distinguish each criterion and places different emphasis on parsimony, that is, the number of free parameters selected in model (Dziak *et al.*, 2014). Generally these criteria favour models that produce a high log likelihood value while using the fewest number of parameters where lower IC values represent better fit. This study will examine the most commonly used IC in LCA, namely the AIC, BIC, the Consistent AIC, (CAIC; (Bozdogan, 1987)) and the adjusted BIC, (aBIC; (Sclove, 1987)).

The AIC is

$$\text{AIC} = -2LL + 2q, \tag{2.3.7}$$

where q represents the number of parameters estimated in the model. Woodroofe (1982) showed that AIC is not consistent and as a result may not select the correct model when sample size is large. Bozdogan (1987) derived a consistent version of AIC to correct for consistency and is

$$\text{CAIC} = -2LL + q\log(n) + q, \tag{2.3.8}$$

where n is the sample size. The BIC is

$$\text{BIC} = -2LL + q\log(n). \tag{2.3.9}$$

BIC has been shown to have a consistent property (Haughton, 1988) which implies that theoretically the true model is more likely to be selected as sample size approaches infinity. Because of this consistent feature of BIC, it may perform poorly when sample sizes are small. Thus, Sclove (1987) proposed an adjustment to the BIC to correct for sample size. The adjusted BIC is

$$aBIC = -2LL + q\log\left(\frac{n+2}{24}\right). \tag{2.3.10}$$

Because of this term difference, BIC penalizes more harshly compared to aBIC for additional parameters included in the model. Model selection is performed by evaluating each IC on each model in consideration. Then the model with the minimum value for that calculated IC is determined as best among the set of candidate models.

2.3.3 Classification Based Information Criteria

Popularized in Gaussian mixture models, classification based information criteria determines the number of latent classes by utilizing the precision of the overall classification assignment. This can be measured by each subject's estimated posterior class probabilities $\hat{\rho_{ik}}$, which specifies subject i's probability of being in each latent class based on the maximum likelihood parameter estimates $\hat{\Psi}$, and observed response pattern $\boldsymbol{u_i}$. Specifically the posterior probability of subject i being in latent class k is

$$\hat{\rho_{ik}} = P(c_i = k|\boldsymbol{u}_i, \hat{\Phi}) = \frac{P(\boldsymbol{u}_i|c_i = k, \hat{\Psi})P(c_i = k)}{P(\boldsymbol{u}_i))}. \quad (2.3.11)$$

A common classification index that utilizes the posterior probabilities is relative entropy (E_k) which measures the overall precision of classification of the whole sample across latent classes (Ramaswamy *et al.*, 1993). E_k is bounded between 0 and 1 where higher values indicate a higher degree of class separation and calculated as

$$E_k = 1 - \frac{E_{k_{raw}}}{n\log(k)}, \quad (2.3.12)$$

where $E_{k_{raw}}$ denotes the raw, unscaled entropy and is defined as

$$E_{k_{raw}} = -\sum_{i=1}^{n}\sum_{k=1}^{K}\rho_{ik}\log(\rho_{ik}) \geq 0. \quad (2.3.13)$$

As noted by Masyn (2013), E_k was never intended for class enumeration because it is possible to have $E_k \approx 1$ yet still have a high degree of error when assigning class membership of certain individuals. Additionally, relative entropy may increase by chance as the number of latent classes increase.

Despite this, classification based information criteria have been proposed that incorporate entropy into the penalty term. Specifically the classification likelihood criterion (CLC) and the integrated classification likelihood criterion (ICL) incorporate the raw, unscaled entropy $E_{k_{raw}}$ into the penalty term. Both CLC and ICL determine the number of clusters in a mixture model rather than the number of components. As argued by Biernacki *et al.* (2000) and Baudry *et al.* (2010), BIC is ideal for determining the number of components in mixture models, but not ideal for determining the *number of clusters* in the data, as it is possible for a cluster to be composed of several mixture components, i.e., the number of components is not taken to be equal as the number of clusters. Therefore, classification based information criterion penalize for both model complexity and how well clusters are separated.

Biernacki and Govaert (1997) established a link between the completed log-likelihood LL_c and the maximized log-likelihood LL, that is,

$$LL_c = LL - E_{k_{raw}}, \tag{2.3.14}$$

where $E_{k_{raw}} \approx 0$ if mixture components are well-separated but if the mixture are poorly separated then $E_{k_{raw}}$ will be a large value. The CLC is defined as

$$CLC = -2LL + 2E_{k_{raw}}. \tag{2.3.15}$$

CLC performs well when the class proportions are constrained to be equal, otherwise, this criterion has a tendency to overestimate the number of classes. This shortcoming was resolved by the ICL which incorporates a heavier penalty term compared to CLC.

Similar to BIC, ICL (Biernacki *et al.* (2000) uses a BIC-like approximation to approximate the integrated log-likelihood value. It essentially derives the ordinary

BIC penalized by entropy

$$\text{ICL} = \text{BIC} + 2E_{k_{raw}}. \tag{2.3.16}$$

We will study these criteria in the context of LCA models and like information criterion, smaller values of CLC and ICL indicate better model fit.

2.4 Overview of Existing Simulation Studies

Many simulations studies have explored the performance of different fit indices under various types of latent variable models and model complexity. There have been a few recommendations in literature but majority point towards the BIC or aBIC as the best information criterion for class enumeration across most modelling conditions (Nylund *et al.*, 2007; Henson *et al.*, 2007; Morovati, 2014; Chen *et al.*, 2017; Tofighi *et al.*, 2008; Whittaker and Miller, 2021). In these studies, aBIC performed better for smaller sample sizes compared to BIC as BIC tended to underestimate the number of latent classes. In studies by Chen *et al.* (2017); Morgan *et al.* (2016); Lukočienė *et al.* (2010), they reported that the CAIC performed as similarly or better than BIC and aBIC. Despite being a frequently reported criterion, AIC performed the worst as it had the tendency to over extract the number of classes (Nylund *et al.*, 2007; Henson *et al.*, 2007).

Studies that compared the BLRT and aVLMR together, concluded that the BLRT showed slightly better performance compared to the aVLMR, but generally the difference was subtle (Morovati, 2014; Nylund *et al.*, 2007). The BLRT was the most reliable fit index across all modelling conditions in Nylund *et al.* (2007) simulation study. But due to its long computational time they advise to only use the BLRT after a set of potential models have been identified. More specifically, Nylund *et al.* (2007) advised to first to assess for potential models with the BIC and then apply the BLRT on these models and compare these results to the BIC. Morgan *et al.* (2016) and Tofighi *et al.* (2008) analyzed the aVLMR, but not the BLRT, and concluded that aVLMR is effective in recovering the number of classes but unperformed compared to the BIC and aBIC.

Morgan *et al.* (2016) and Morgan (2015) considered the ICL on mixed mode LCA models and latent profile analysis models, respectively. Recommendations from these studies indicated mixed results for the ICL. As well, Henson *et al.* (2007) studied ICL and CLC on structural equation models and noted that ICL and CLC performed very well compared to other fit indices in certain conditions but overall showed limited utility. Alternatively, McLachlan and Ng (2000) reported the results of three simulation studies and found that of all the criteria considered, ICL and CLC were the only criteria to correctly select the true number of classes in all three studies.

In another study, Fonseca and Cardoso (2007) examined ICL and CLC along with multiple information criterion on mixture models of categorical, continuous and of mixed data type. The ICL was best for mixed data type as it achieved accuracy rates of 80% followed by CAIC and BIC, both with 70%. The aBIC, CAIC, BIC and ICL proved to be highly reliable for identifying the correct number of latent classes in Diallo *et al.* (2017b) but recommended against the the AIC and CLC.

Within the enumerating process with covariates, there is debate on whether the inclusion of covariates could improve class enumeration accuracy. This is a concern because the number of classes and thus how they differ from one another may change drastically when covariates are included in the enumeration process. Nylund-Gibson and Masyn (2016) analyzed the impact of misspecifying covariate effects during the class enumeration procedure by analyzing the enumeration accuracy on five different LCA models with distinguishing covariate pathways. They found that class enumeration is best with an unconditional model and advised for no covariates. They showed that an incorrect and misspecfied covariate effect would be more detrimental on the performance of the model compared to using an unconditional model for class enumeration.

Collins and Lanza (2010) and Diallo *et al.* (2017a) also advocate that covariates should only be added to the modelling process only after the enumeration process is done with an unconditional model. When covariates were included results were uniformly poor and hampered the class enumeration accuracy in Tofighi *et al.* (2008) class enumeration study on growth mixture models. Enumerating without covariates is also the recommended way to begin the three step method which currently viewed as the best way to include covariates into the modelling process (Nylund-Gibson and Choi, 2018).

Those that argue for the inclusion of covariates observe that including covariates simultaneously during the enumeration process increase model accuracy. Li and Hser (2011) study indicated that incorporating covariates during the enumeration process would help more replications converge to the correct number of classes. Lubke and Muthén (2007) examined this problem in the context of factor mixture models and suggested that including covariates during the class enumeration process provides additional knowledge which may increase the performance of the fit indices in certain cases such as when class separation and homogeneity is low. Peugh and Fan (2015) concluded that the decision to incorporate covariates vary depending on the modelling situation. They showed that the inclusion of covariates had negligible effect for small sample sizes but was beneficial for larger sample sizes with underlying conditions such as poor class separation or proportion conditions.

Multiple simulation studies have been produced through recent years but taken as a whole, it is difficult to conclude which fit index is more preferred for class enumeration. Similarly determining whether the addition of covariates has detrimental or beneficial impact on enumeration accuracy is also unclear. Thus this simulation study provides more insight into these issues.

Chapter 3

Simulation Study

This simulation study focuses on latent class regression models with the aim to: 1. quantify the accuracy of commonly used fit indexes in the context of class enumeration. 2. determine whether the inclusion of covariates can improve class enumeration. 3. highlight factors that may lead to better or worse performance of latent class regression models. We use the `Mplus Automation` package Hallquist and Wiley (2018) to assist with organization and creation of input files of this study. It is an `R` package R Core Team (2020) that leverages `Mplus` (Muthén and Muthén, 2017) to facilitate with large, complex simulation studies and run large-scale latent variable models.

3.1 Population Models

All data were simulated from two population models with known covariate effects and analyzed under alternative models that vary in the number of classes and covariate relationships. The level of model complexity is also studied as it may play in the decision to include covariates in the extraction process. Therefore in this study we analyze two population models of varying level model complexity. Specifically, population A is a latent class model with an indirect effect as shown in Figure 3.2a. Population B extends population A by incorporating direct effects as shown in Figure 3.2b. We refer to population A as the 'simple model' because it only has one indirect pathway and the local independence assumption is not violated. Population B is the 'complex model' and illustrates a more complex interrelationship between covariates, items and the latent variable. Furthermore, the impact on latent class membership is indirectly impacted by x as well as a secondary covariate w. In this case, the local independence assumption is relaxed by incorporating a direct effect between u_1 and u_2. Population B is based on a similar model used in the medical diagnostic literature as described in Qu *et al.* (1996).

Population parameters were chosen to be reasonably representative of LCA models in applied research. We based our population design on Morgan (2015) empirical condition review on applications of LCA. It was observed that 3-class solutions and the use of binary items were the most frequent in applied studies. In this simulation study, we consider three-class LCR models with ten binary items which were analyzed at four sample sizes $n = 200, 500, 1000$ and 2000. Previous studies indicate that this range of sample sizes were the most commonly used and expected to function adequately for detection of latent classes (Nylund-Gibson and Choi, 2018). Other factors that we manipulate to vary the the distinctness of the latent classes, include the class proportions, item-response probabilities, and the magnitude of the covariate effect.

We consider two different splits for the class membership probabilities. Split 1 corresponds to a poorer class split where there is one large or normative class, a moderate class and smaller or rarer class: $\pi_1 = 0.60, \pi_2 = 0.35, \pi_3 = 0.05$. Split 2 examines two classes of similar size and a relatively smaller class: $\pi_1 = .40, \pi_2 = .45, \pi_3 = .15$. The level of class separation is an obvious factor and thus will be thoroughly studied. Several studies have shown an equal classes case, however, (Morgan, 2015) review indicated that equal class proportions are rarely observed in empirical LCA applications. Based on (2.2.3), the class proportions are represented by the multinomial regression intercept parameters γ_{0k}. In particular we specified $\gamma_{0k} = 2.70, 2.16, 0$ for split 1 and $\gamma_{0k} = 1.09, 0.90, 0$ for split 2, for $k = 1, 2, 3$.

The distinctness of latent classes is also impacted by the overall quality of latent class items. Generally if the model contains several "good" items then we would expect high class enumeration accuracy. A "good" item is one that can measure the latent variable reasonably well allowing for better interpretation of the latent classes. This strong item-class relationship is characterized by having both a high degree of class homogeneity and class separation (Masyn, 2013). Class homogeneity specifies whether there is a specific response category on that item that strongly characterizes that latent class. A specific threshold for high class homogeneity varies but generally item-response probabilities greater than 0.7 or less than 0.3 is accepted (Masyn, 2013). Class separation defines how distinguishable the latent classes are. Particularly, when there is high degree of class separation, the overall response pattern is uniquely characteristic to that particular latent class only and not for any other classes (Collins and Lanza, 2010). In our study we examine how the quality of items impact class enumeration accuracy of each fit index.

The degree of homogeneity and separation were varied by manipulating the item-response probabilities to create three different design conditions: high, moderate and low quality conditions, as shown in Table 3.1. These conditions were similar to the ones used in (Yang and Yang, 2007) simulation study on LCA models. Note that in `Mplus`, the item-response probabilities are translated as thresholds τ_{mk} and calculated according to (2.2.6). The calculated threshold values used to simulate each design condition can be found in Appendix A. The design of each condition is similar in structure but the bolded values dictate the degree of class homogeneity and separation. The items related to the bolded values characterizes each latent class of endorsing a particular item. For example, latent class one would be interpreted according to items 1,2,3, whereas latent class two would be interpreted according to items 4,5,6. Therefore, we say items 1,2,3 characterize latent class one and items 4,5,6 characterize latent class two. This interpretation is equivalent to the concept of factor loadings used in factor analysis where high factor loadings on a set of items define a clear factor.

The first design captures an example of high class homogeneity and separation in which associated with each class there is a definite response pattern that is much more likely expressed compared to other response patterns. For instance, individuals in latent class one are likely to endorse items 1,2,3 with a high probability of 0.90, i.e., an estimated 90% of individuals will endorse these items and 10% will not. Likewise individuals in latent class two and three endorse items 4,5,6 and items 7,8,9 with high probability, respectively. Therefore in each latent class there is a highly distinguishable response pattern that characterizes that class.

The second design is an example of moderate class homogeneity and class separation. Classes are still relatively distinguishable and the same items characterize each class, however, the probability of endorsement decreased, ranging between 75%

Table 3.1: Item-response probabilities used for data simulation. The interpretation of the latent classes is determined by the overall pattern of item-response probabilities. The bolded values indicate which items characterizes each latent class in the high and moderate case. However in the low case, these same items are no longer characteristic of each latent class because there are no items that clearly distinguishes each latent class.

	High			**Moderate**			**Low**		
Items	Class 1	Class 2	Class 3	Class 1	Class 2	Class 3	Class 1	Class 2	Class 3
1	**0.9**	0.1	0.15	**0.75**	0.3	0.15	**0.5**	0.45	0.15
2	**0.9**	0.2	0.15	**0.75**	0.15	0.15	**0.5**	0.4	0.15
3	**0.9**	0.25	0.3	**0.75**	0.35	0.3	**0.5**	0.3	0.3
4	0.5	**0.9**	0.3	0.5	**0.8**	0.3	0.5	**0.45**	0.3
5	0.1	**0.9**	0.2	0.1	**0.8**	0.2	0.4	**0.45**	0.35
6	0.5	**0.9**	0.25	0.5	**0.8**	0.25	0.5	**0.45**	0.25
7	0.1	0.25	**0.9**	0.1	0.3	**0.75**	0.35	0.3	**0.5**
8	0.1	0.2	**0.9**	0.1	0.4	**0.75**	0.45	0.4	**0.5**
9	0.35	0.2	**0.9**	0.35	0.2	**0.75**	0.2	0.45	**0.5**
10	0.35	0.25	0.2	0.35	0.4	0.45	0.3	0.4	0.45

- 80%. Therefore, we can still equate items 1,2,3 as characteristic of latent class one, 4,5,6 as characteristic of latent class two and 7,8,9 characteristic of latent class three but compared to the first design, we are less confident in our interpretation. By decreasing the item response probabilities we are effectively decreasing the quality of the items and increases the difficulty in identifying a characteristic response pattern for each class.

The last design depicts an example of low class homogeneity and separation. The same bolded items are no longer characteristic of each latent class. For example, probabilities of items 1,2,3 and items 4,5,6 are almost indistinguishable in latent class one and two, so there is a high level of uncertainty between these classes. In other words, the difference between these two classes are not as pronounced compared to the previous two cases because the probability of expressing a certain response pattern in class one is also likely to be expressed in class two. Generally latent classes are more distinguishable when there are particularly high (or low) probabilities for a given class. In the low quality condition, there are no substantial differences in the overall item-response probabilities.

The relationship between class homogeneity and class separation is more transparent when the item-response probabilities from Table 3.1 are plotted, as shown in Figure 3.1. It is much easier to visualize the qualitative differences of each latent class and evaluate the degree of class homogeneity and separation. In the first and second panel, interpreting the latent classes would be straightforward because the item response probabilities across the measured items clearly differentiate the latent classes. In contrast to the last panel, interpreting the latent classes is no longer as straightforward but more difficult because of the of low class homogeneity and separation.

As a whole, we can see that a high degree of class separation guarantees a high degree of class homogeneity. However, it is important to note that the converse is not true, namely a high degree of class homogeneity does not guarantee high class separation. For example, consider a 2-class model, one class endorses a particular item with probability of 0.95 and the other class endorses the exact item with a probability of 0.90. In this case, class homogeneity is high but classes are indistinguishable because that item is highly characteristic in both classes. Alternatively consider a 2-class model where the probability of endorsing a particular item is 0.85 in class one and 0.10 in class two. In this case, we have a high degree of class separation as well as class homogeneity such that the first class may be interpreted as being more characteristic endorsing that item whereas the second class will not.

Figure 3.2 summarizes the covariate relationships considered in the study. We manipulate the strength of the covariate effect to determine whether or not it influences the class enumeration accuracy. For population A, a continuous covariate x was generated from a standard normal distribution in which x has an indirect effect on the items via the latent variable. The effect of x on c is measured by the regression parameter γ_{1k} in (2.2.3). In particular, $\gamma_{1k} \in 0.40, 0.90, 1.50$ to represent small, moderate and large effect sizes which correspond to Cohen's d of 0.20, 0.50 and 0.80, respectively (Cohen, 1988).

In population B, w is a binary categorical covariate and has a direct effect on u_1 and u_2. In this case, the local independence assumption is violated because w is a common cause beyond the latent variable. x is a covariate that follows a standard normal distribution and has a moderate effect on the items through c, i.e., the regression parameter γ_{1k} is fixed at 0.90. Using (2.2.4), population B is defined as

$$P(\boldsymbol{u}_i|w_i, x_i) = \sum_{k=1}^{3} \left[P(c_i = k|x_i) P(u_{1i}, u_{2i}|c_i = k, w_i) \prod_{m=3}^{10} P(u_{mi}|c_i = k) \right]. \quad (3.1.1)$$

The latent variable c and w account for the shared association among u_1 and u_2, thus, u_1 and u_2 are conditionally independent with respect to c and w

$$P(u_{1i}, u_{2i}|c_i = k, w_i) = P(u_{1i}|c_i = k, w_i) P(u_{2i}|c_i = k, w_i), \quad (3.1.2)$$

where the probability of endorsing item u_1 or u_2 is given by

$$P(u_{ji} = 1|c_i = k, w_i) = \frac{1}{1 + \exp(\tau_{u_{jk}} - \beta_{u_j} w_i)} \quad j{=}1{,}2. \quad (3.1.3)$$

The direct effects of w on u_1 and u_2 are equal such that $\beta_{u_1} = \beta_{u_2}$. The impact of direct effects on class enumeration are studied for small, moderate and large effect sizes which correspond to $\beta_{u_j} = 0.40, 0.90$ and 1.50, respectively, for $j = 1, 2$.

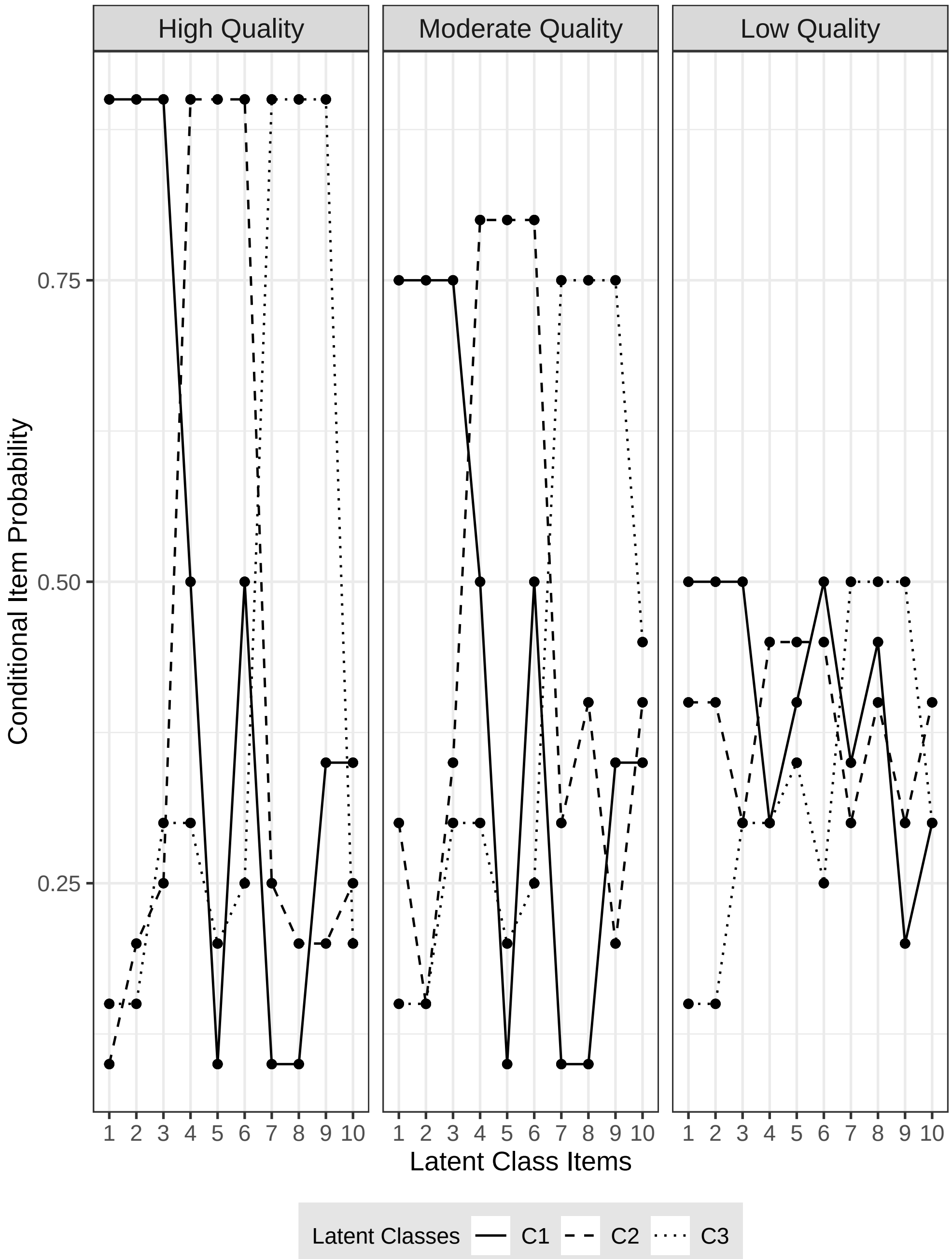

Figure 3.1: Interpretation of latent classes is based on the overall distribution of the item-response probabilities. Each panel represent a different degree of class homogeneity and separation. In general as class separation decreases so does class homogeneity but the converse is not true.

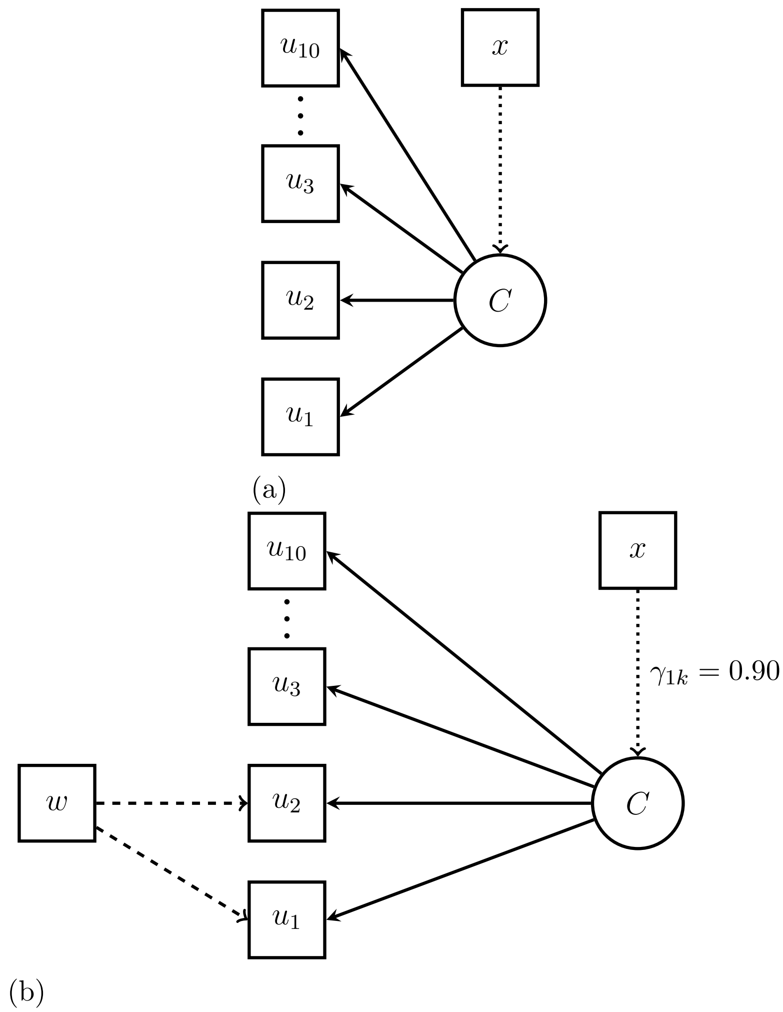

Figure 3.2: For population A, the covariate x has an indirect effect on items via the latent class variable. For population B, two covariates are simulated. The covariate x has an indirect effect on items which is fixed at $\gamma_{1k} = 0.90$. w has a direct effect on u_1 and u_2.

3.2 The Analysis Procedure

The simulation design consist of two population models, four sample sizes, three sets of conditional probabilities with two class splits and three covariate effect sizes. The factors manipulated resulted in a total of $2 \times 4 \times 3 \times 2 \times 3 = 144$ unique design conditions. 500 replicated data sets were simulated according to each design condition and analyzed with different LCA models that ranged from two through five latent classes using `Mplus` (Muthén and Muthén, 2017). Results of each fit index from every run were collected from output files, and saved for analysis with `R` (R Core Team, 2020) and `Mplus Automation` (Hallquist and Wiley, 2018).

Additionally to determine whether the inclusion of covariates improved the ability to detect the correct number of latent classes, each replication was analyzed twice: once with covariates, and once without covariates. When covariates were included, the specification for these runs matched the population models according to where the replication was simulated from. Our outcome of interest is to determine how frequent the correct number of latent classes could be identified by each information criterion and likelihood ratio test. To capture this, we recorded the proportion of replications (rounded to the whole number) that each fit index selected the two, three, four, and five-class model as the best solution under each condition.

Results obtained in this study were not based on local solutions. Before initializing the study, multiple sets of random starting values were assessed on both population models to ensure a reasonable number of starting values would consistently converge to the same global solution. We opted for 150 sets of random starting values for the initial stage of the optimization and 50 starting values for the final stage of the optimization process. However, there were a number of replications that were problematic, specifically the 5-class runs. A common way to deal with non-convergence is to discard these replications and based the results on replications where all the class runs fully converged (i.e., all 2,3,4 and 5-class run properly converged).

Rather than retaining replications with only full class solutions, we proceed with another option that is more realistic when a large number of non-convergence class runs are encountered in latent variable applications. This was also the procedure Tofighi *et al.* (2008) applied in their study. Rather than removing the replication, we remove the run that failed to converge and assume the previous run that converged as the class solution. In other words, if the k-class model failed to converge for a replication, it is reasonable to take the previous (k-1)-class run as the class solution. Therefore when the 5-class models failed to converge, we selected the 4-class model as the correct solution, provided the fit index supports the 4-class model. Of course if a given fit index unequivocally selects 2-class or 3-class model then the non-convergence of the 5-class model is unimportant.

The aVLMR and BLRT can be requested in `Mplus` with `Tech11` and `Tech14`, respectively. Recall the BLRT and aVLMR are nested test where the (k-1)-class model is obtained by restricting the parameters of the first latent class of the k-class model. Multiple starting values were used to estimate the best (k-1)-class model in which 150 sets of random starting values were selected for the first step of the optimization and 50 sets of starting values were selected for the second step of the optimization. To ensure we were achieving reliable p-values for the BLRT, we used 100 bootstrap samples as suggested by study McLachlan (1987), which noted that at least 99 bootstrap samples were necessary to obtain optimal power. (Dziak *et al.*,

2014) also used 100 bootstrap sample in their BLRT power analysis. To ensure no local maximum occurred during the analysis of the bootstrap samples, we specify random starting values for (k-1) and k-class model with `LRTSTARTS=0 0 150 50`.

Class enumeration results relating to aVLMR and BLRT can be found in Appendix B with reference to Tables B.1, B.2, B.3 and B.4. The tables summarize the proportion of replications that each latent class model was selected based on aVLMR and BLRT. In this case, two- through five- latent class models were considered for each replication and provided a p-value that compare the fit of a latent class model with (k-1)-classes to one with k-classes. Note that when a two-class model is specified, a one-class model is also considered thus we have a total of five possible class solutions in our study. The number of classes is determined based on how well each LRT can differentiate neighbouring class models that differ by one class from the true population model (Nylund *et al.*, 2007). More specifically to earn support for a k-class model we require a significant p-value for the (k-1)-class model during the k-class run and a non-significant p-value for the k-class model during (k+1)-class run. If this is not satisfied during the (k+1)-class run, we continue to add classes until the first non-significant p-value is observed.

Information criteria were studied in a similar fashion such that a sequence of latent class models ranging from two- through five- classes were fit on each replication. These models were then compared and scored based on each IC, then selected based on where the minimum value occurred across each criterion and model. Table B.9 through Table B.16 summarized the proportion of replications each latent class model was identified as correct according to each IC. Classification based IC followed an identical process as the IC, in which model selection was based on the where the lowest value occurred across the models considered. Table B.5, B.6, B.7 and B.8 showcase the proportion of replications a particular latent class model was selected based on the ICL and CLC.

Another interest was to study how certain design factors contribute to the class enumeration accuracy. The performance was analyzed graphically by averaging the proportions of times each fit index selected the true three-class solution. Specifically, each graph is created to display the two factors simultaneously with the other factors averaged out. For example, Figure 4.1 shows the impact of sample sizes and quality of the item-response probabilities on each fit criterion, which is calculated by averaging the proportion of replications that selected the three-class model over all the covariate effect sizes and class splits. Likewise, Figure 4.2 displays the impact of different covariate effect sizes and class splits by averaging over the all the sample sizes and quality of item-response probabilities.

Chapter 4

Results

In this section, we present the results of the simulation study, in particular the proportion of replications that selected the correct number of classes for each fit index. This includes comparing the performance of each one under different modelling conditions and determine whether the inclusion of covariates is reflected in the class enumeration accuracy. The full simulation results are given in Appendix B. Additionally coverage values of 95% confidence intervals are discussed.

4.1 Average Coverage Values

The average coverage values across all parameter estimates were calculated. For a simulation study to be meaningful, the parameter estimates obtained from analysis must resemble the population parameters. The precision of interval estimates can be measured by coverage values. Coverage values state the proportion of replications for which the 95% confidence intervals contain the true population parameter Muthén and Muthén (2002). The coverage of the estimates obtained from our study were good and fell in 91% to 97% range. However, coverage values of the smaller latent classes were low, particularly when sample size is n=200. This is unsurprising given that 5% of 200 only produces a class size of 10. Likewise 15% of 200 corresponds to a class size of 30. Recall we considered two sets of class proportions in which the smallest latent class was 5% and 15% in split 1 and 2, respectively.

The average coverage estimates of the smallest latent class of each spit is shown in Table 4.1. It illustrates the impact of sample size and quality of items on coverage. Coverage values improve as the quality of items improve and sample size increases. For instance in population A with low quality item-response probabilities, coverage is 68% at n=200, 81% at n=500, 85% at n=1000, and 91% at n=2000. In contrast, coverage is improved drastically with high quality item-response probabilities such that coverage increased to 92% at n=500.

Table 4.1: Coverage estimates were very good for the larger latent classes, but poor for the smaller latent classes. Coverage estimates increased as sample size and quality of item-response probabilities improved.

		Population A		Population B	
Quality	Sample Size	Class Size of 5% Coverage Values	Class Size of 15% Coverage Values	Class Size of 5% Coverage Values	Class Size of 15% Coverage Values
Low Quality	200	68	84	71	77
	500	81	89	82	85
	1000	85	90	88	88
	2000	91	93	92	91
Moderate Quality	200	71	87	76	90
	500	93	94	91	90
	1000	96	95	93	95
	2000	95	96	94	96
High Quality	200	76	93	78	91
	500	92	95	93	92
	1000	94	95	94	97
	2000	96	97	93	97

4.2 Sample Size

Figure 4.1 gives the average proportion of replications in which a three-class model was correctly selected when sample size varied. Overall, we see that majority of the fit index experience increase accuracy rates as sample size increase. For example, consider population A, row moderate quality, we see that as sample size increases from left to right, the accuracy rates of the fit indices shift upwards. Additionally as sample size reach n=1000, with the exception of AIC, CLC and ICL, all the fit indices achieved enumeration accuracy rates greater than 90% and near perfect when sample size approach n=2000. This increasing trend in the accuracy rates is also portrayed in population B.

The aBIC, BIC and BLRT performed similarly well and enumerated the most accurately across different sample sizes and population models. For n=200, results were mixed and varied depending on the quality of item-response probabilities. AIC performed the worst of the information criterion and the enumeration accuracy did not reflect any improvement with increase sample size, in fact it appears that accuracy rate decreased as sample size was increasing. Similar trends were exhibited for the ICL but the change in accuracy rate decreased much more drastically such that accuracy is less than 10% when sample size reach n=1000 and close to 0% for n=2000. The CLC was the worst performing fit index where enumeration rate was near 0% across all sample sizes.

4.3 Quality of Item-Response Probabilities

Three different structure of item-response probabilities were examined in this simulation study. Figure 4.1 summarizes the average enumeration accuracy of selecting the three class model under high, moderate and low quality conditions. It was apparent that the overall quality of the item-response probabilities had a bearing on class enumeration. The proportion of replications that correctly identified the three class model decreased as the quality of the item-response probabilities decreased. Specifically moving downwards from high to low quality conditions in Figure 4.1, the average proportions of correct enumeration drops substantially across both population models. For example with high quality items, the aBIC, BIC, CAIC achieved 90% accuracy rates with a sample size of n=500, however, given moderate quality conditions it required at a sample size of at least n=1000.

With high quality conditions, majority of the fit indices, with the exception of AIC, ICL and CLC, gave comparable performance and achieved high enumeration accuracy rates ranging between 80% to 100%. With moderate conditions the difference in performance of each fit indices can be seen more clearly. The aBIC and BLRT performed comparably and identified the highest number of three class solutions. Given low quality conditions, the accuracy rates were very poor and class solutions were barely identifiable by most of the fit indices. However, the AIC and ICL had some enumeration capability though accuracy rates were generally less than 40%. As a whole, Figure 4.1 showcase the relationship between quality of items and sample size and emphasizes not only the importance of obtaining or deriving "good" items during the data collection process but utilizing a reasonable sample size as well.

4.4 Class Proportions

Two different sets of class proportions were considered in the simulation study where split 1 represent a large, moderate and rare class split (60-35-5) and split 2 simulated two classes of similar size and one smaller class (45-40-15). It is evident that the overall class split had an impact on class enumeration as indicated by the spread across each split in Figure 4.2. The accuracy rates of each fit index is much more spread in split 1 compared to split 2, which implies that enumeration accuracy is higher when latent classes are more equal in size. This is anticipated because even with a relatively large sample size such as n=2000, 5% of 2000 produce a class size of 100 may be too small of a class to be detected by certain fit indices. In the case of dissimilar class sizes such as in split 1, it appears that the aBIC and BLRT would be more successful at detecting smaller/rarer classes in the population.

4.5 Model Complexity and Effect Size

Figure 4.2 captures the average number of replications that selected the three-class model given different levels of effect sizes. In general the effect size contributed from a covariate does impact the class enumeration accuracy. To illustrate, consider population A, split 1 with a small effect, the accuracy of the fit indices, with exception of the ICL and CLC, range between 37% and 60%. As we move down the plot to a larger covariate effect size, majority of these fit indices experience an increase in the number of identifiable class solutions. The AIC, ICL and CLC have no changes in accuracy as the effect size increases. This is the general pattern expressed across all the modelling conditions considered in Figure 4.2.

Model complexity also had an influence on the class enumeration accuracy. Population A and B were very similar in construction with three latent classes, ten binary items and an indirect effect specification onto the latent variable. The difference was population B which incorporated an additional direct effect specification between item u_1 and item u_2. By specifying the direct effect between items, we relaxed the local independence assumption and increased the number of free parameters to be estimated, thus making population B slightly more complex.

Generally fit indices that performed well in population A also performed well in population B with minor differences depending on the quality of the item response probabilities. With high quality conditions, the difference in accuracy rate between population A and population B were minimal as seen in Figure 4.1. The difference is more prominent when the quality of the item-response probabilities decrease. With moderate quality conditions, accuracy rates of population B are slightly worse compared to population A. This trend is also exhibited in Figure 4.2. The difference in accuracy rate between the two population models can be more clearly seen in Table 4.2 which summarizes the average proportion of replications that correctly selected the three class model according to each of the fit indexes, categorized by sample size.

4.6 Including Covariates

Determining whether the inclusion of covariates could improve the performance of an fit index required each replication to be conducted twice: with and without covariates. When covariates were included, the specification matched the population models based on where the replication was simulated from. Referring back to Figures 4.1 and 4.2, these figures showcase the average proportions of replications that selected the three class model according to each fit index, with and without covariates. An upward slope signified that including covariates improved the accuracy rate of a particular fit index. Overall the inclusion of covariates proved beneficial for class enumeration across both populations models.

Incorporating covariates improved the enumeration accuracy under certain troublesome conditions such as poor item-response probabilities or sample size issues. To illustrate, consider the case of population A, moderate quality conditions, and sample size of n=500 in Figure 4.1, accuracy rates of the aBIC and BLRT saw an improvement of roughly 8% when covariates were included. Similar pattern of results were observed for aVLMR, BIC and CAIC. Including covariates when the effect size was small or when the class proportions were imbalanced, dissimilar in size such as the case of split 1 also demonstrated improvement in enumeration accuracy. As shown in Figure 4.2, the aBIC, BIC, CAIC, aVLMR and ICL all experienced improvements with covariates.

This improvement is not surprising since the conditional model was specified to match the population model according to where the replication was simulated from. But in most applications of LCA with covariates, the true covariate specification is unknown. One could hypothesize what the covariate relationships may be but knowing the exact and correct specification is unlikely. In a simulation study by Nylund-Gibson and Masyn (2016), they showed that incorporating misspecified covariate effects into LCA models early on in the enumeration process would commonly lead to an over extraction of latent classes.

Additionally we note that the magnitude of improvement between the conditional and unconditional model was minimal and usually unsubstantial. In other words, the model without covariates does just as well as the the conditional model with true covariate specification. This is noteworthy because conducting the enumeration process without covariates saves a great amount of computational time. For instance with population B, estimating a 3-class run with covariates took approximately 76 seconds to compute in comparison to approximately 8 seconds without covariates. In our case the populations models were relatively simple and start values considered were much smaller compared to values used in real settings. With larger, more complex models and larger range of starting values the difference in computation time can range from a few minutes to hours (Dziak *et al.*, 2014; Asparouhov and Muthén, 2012), assuming the same fit indices are used as in this study.

Therefore, we recommend that it is reasonable and best to first conduct class enumeration with an unconditional LCA model to determine the number of classes. Once the number of classes is finalized and there is evidence of indirect/direct effects in the model, covariates should then be included into the analysis.

4.7 Extraction Errors

The results collected so far provide the proportion of times each fit index correctly selected the number of latent classes across different modelling conditions. We would also like to understand the type of errors committed when an incorrect model is selected (i.e., extracting too many or too few classes). Table 4.2 summarizes the average proportions of replications that were incorrectly and correctly selected by each fit index, categorized by sample size.

The AIC had a tendency to overestimate the the number of latent classes and the error became more prominent as sample size increased. These results are unsurprising given that the penalty term of the AIC generally favoured models of more complexity and its inconsistency property (Woodroofe, 1982). The BIC, CAIC and aBIC had a tendency to underestimate the number of latent classes and because these IC are consistent (Dziak *et al.*, 2018), they enumerate the correct the number of latent classes more frequently as sample size increased.

BIC, CAIC and aBIC penalizes more harshly compared to the AIC and as a result they favour smaller and simpler models. Likewise, the aVLMR and BLRT typically underestimated the number of latent classes. The ICL and CLC had the largest extraction errors and frequently overestimated the number of latent classes. The error reached over 40% as sample size increased to n=2000. In our simulation study, when a fit index went wrong it would typically underestimate the number of latent classes.

Table 4.2: Average proportion of replications correctly selected as well as the average proportion of replications that resulted in either under or over extracting the number of latent classes according to each fit index by sample size.

		Population A			Population B		
Sample Size	Criterion	Under extraction	Correct Identification [a]	Over extraction	Under extraction	Correct Identification [a]	Over extraction
200	AIC	10.1	37.6	20.8	10.3	38.5	20.8
	BIC	28.5	**42.8**	0.0	28.7	41.5	0.0
	aBIC	12.8	**42.5**	14.3	13.0	**41.6**	14.1
	CAIC	31.7	36.4	0.0	32.4	35.3	0.0
	aVLMR	26.2	42.4	2.6	29.3	36.7	2.4
	BLRT	26.4	40.1	3.2	20.2	**45.8**	3.9
	ICL	17.3	**51.7**	6.9	17.6	**46.3**	7.7
	CLC	0.1	1.8	48.9	0.2	2.2	48.7
500	AIC	7.9	36.7	23.5	7.8	36.7	23.3
	BIC	21.5	**57.1**	0.0	21.8	**56.0**	0.0
	aBIC	17.1	**61.5**	0.5	16.7	**60.8**	0.5
	CAIC	22.5	55.1	0.0	23.5	52.7	0.0
	aVLMR	20.0	54.9	2.4	20.1	54.3	3.3
	BLRT	17.8	**57.4**	2.8	17.4	**57.2**	2.8
	ICL	4.1	28.0	31.9	4.7	30.2	30.2
	CLC	0.3	3.3	48.0	0.5	3.4	47.8
1000	AIC	6.2	37.3	24.4	6.2	35.4	25.4
	BIC	17.9	**63.3**	0.0	17.1	**61.6**	0.0
	aBIC	16.7	**65.2**	0.0	16.6	**64.0**	0.0
	CAIC	18.7	61.6	0.0	17.9	58.6	0.0
	aVLMR	17.5	59.3	2.7	18.3	58.8	3.8
	BLRT	15.9	**63.0**	2.2	15.6	**61.8**	2.4
	ICL	1.7	14.9	40.9	1.9	15.2	40.5
	CLC	0.5	3.1	47.9	0.5	3.3	47.9
2000	AIC	5.1	35.9	26.1	4.4	35.2	27.9
	BIC	16.7	**66.6**	0.0	16.7	**64.5**	0.0
	aBIC	16.6	**66.4**	0.0	16.7	**66.7**	0.0
	CAIC	16.7	64.5	0.0	16.7	64.2	0.0
	aVLMR	16.0	61.8	3.1	16.1	61.7	3.0
	BLRT	15.5	**65.4**	2.2	15.0	**66.1**	2.0
	ICL	0.9	7.7	45.3	0.8	8.5	45.0
	CLC	0.4	3.2	48.0	0.3	3.4	48.0

[a] Bolded values indicate the top 3 best performing index in each condition. Averages were calculated across different class splits, quality conditions and effect size of each population model.

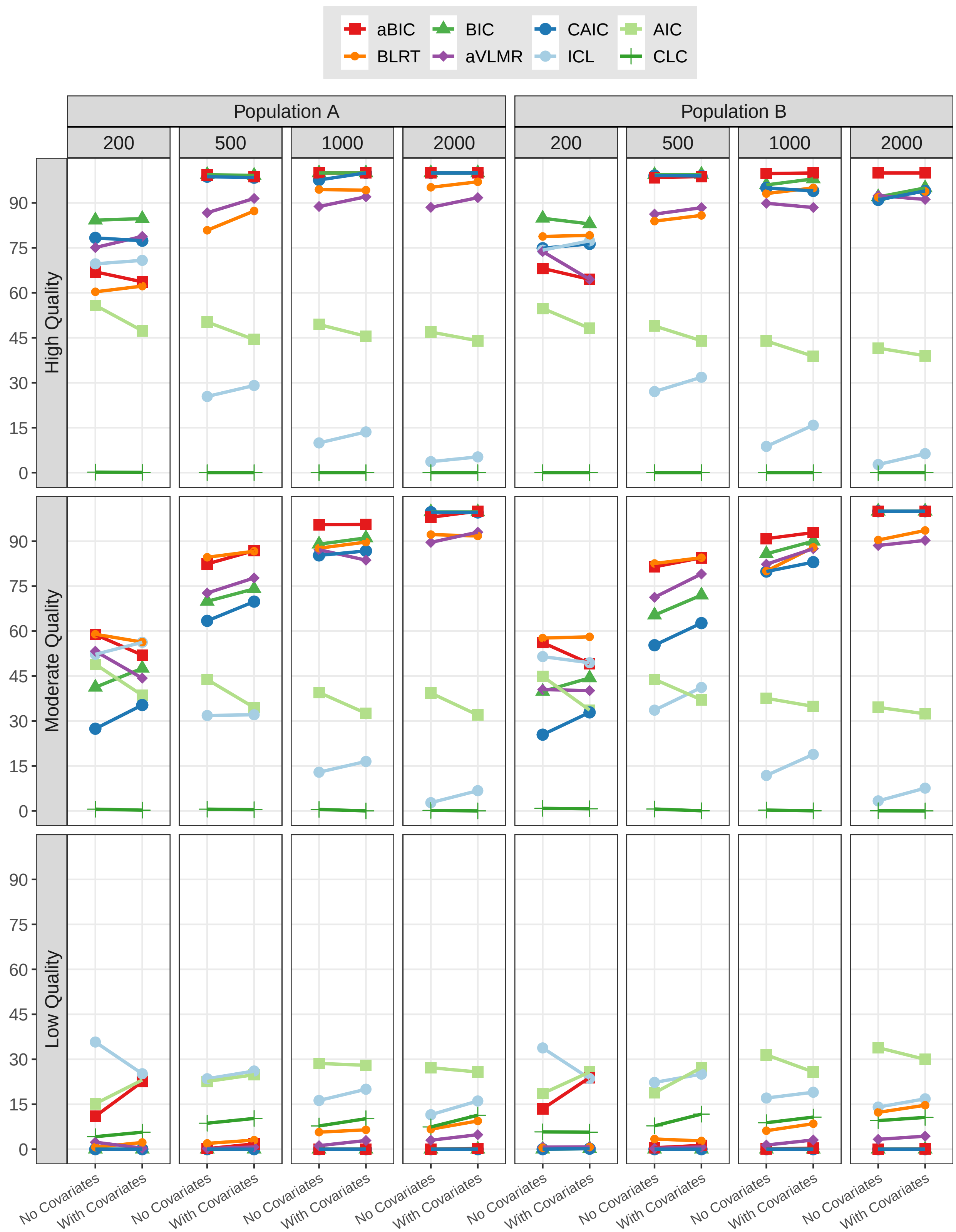

Figure 4.1: Average proportion of replications that selected the 3-class model based on Quality of Indicators and Sample Size.

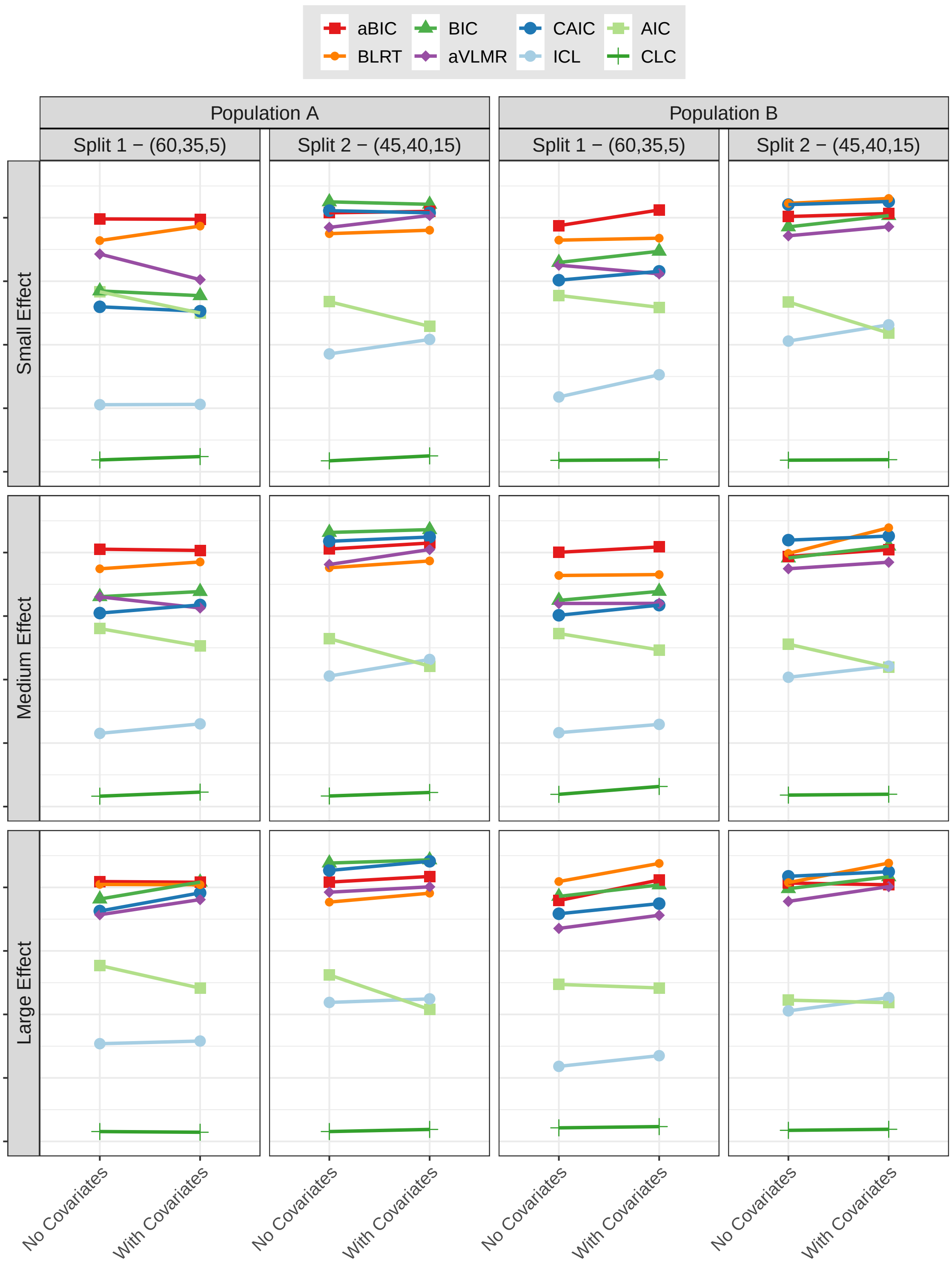

Figure 4.2: Average proportion of replications that selected the 3-class model by Covariate Effect Size and Class Proportion Split.

Chapter 5

Discussion

We conducted a simulation study that examined the performance of several commonly used fit indices in LCA applications. The main objective was to better understand which of these fit indices were more preferable as a model selection tool for LCR models, as well as provide more insight as to whether class enumeration should be conducted with or without covariates. Lastly, we studied how certain design factors would contribute to the class enumeration accuracy and highlighted which would lead to better or worse performance.

The shaded column of Table 4.2 summarizes the average proportions of replications that correctly identified the three-class solution for each fit index, grouped by sample size. These averages were calculated across different class splits, quality conditions, and effect size of each population model. The bolded values of this column indicate the top 3 best performing fit indices. Overall, the aBIC and BLRT selected the correct number of latent classes most frequently across all modelling conditions. Generally as sample size increased, the performance of the BIC, CAIC and aVLMR were comparable to the aBIC and BLRT. The CLC was the worst performing fit index across all modelling conditions. The ICL should not be dismissed entirely because the ICL performed with the highest accuracy for n=200, at 51% and 46% for population A and B respectively. More analysis is needed to determine whether this was due to chance or the ICL can be used as a class enumeration tool for small sample sizes.

Between the two likelihood ratio tests, the BLRT showed better performance than aVLMR but accuracy rates between the two were quite close for majority of the modelling conditions considered in this study. Nylund *et al.* (2007) study found similar results but noted that the p-value for aVLMR tended to bounce between significant and non-significant as the number of latent classes increased. In comparison to the BLRT, once the BLRT p-value was non-significant it remained so for subsequent increased class models (Nylund *et al.*, 2007). Thus, they recommend that when the first non-significant aVLMR p-value occurs it is a good indication to stop increasing the number of classes. We did not encounter this problem in our study. Though the BLRT performed well it is limited due to its computation time. Therefore, similar to the recommendation of Nylund *et al.* (2007), class enumeration should first be conducted with the aBIC or BIC to determine a set of potential candidate models. Once that is determined, the BLRT can enumerate on the set of candidate models and confirm whether results obtained by the BLRT match results from the aBIC and BIC.

In terms of the information criteria, findings in this study suggest that the AIC should not be used a model selection tool as it tended to overestimate the number of latent classes. The CAIC and BIC performed very similarly to each other in most conditions. But given their penalty terms this is not a huge surprise since CAIC is the BIC with an added term for the number of parameters, thus is penalizes slightly more harshly compared to the CAIC and favour simpler models more than the BIC. Between the BIC and aBIC, the aBIC performed slightly better across our modelling settings.

This study varied several design factors including the sample size, quality of item-response probabilities, model complexity, class proportion split and size of the covariate effect. Unsurprisingly items with high class separation and class homogeneity contributed to how well fit indices were able to distinguish the latent classes, as illustrated in Figure 4.1. When class separation and class homogeneity decreases so does the accuracy rate of the fit indices. In some cases, increasing sample size or improving the quality of items could compensate for less ideal conditions. For example, moderate quality conditions can be compensated by increasing the sample size and high quality conditions compensating for small sample sizes.

The overall class proportion split and size of each of the latent class also had bearing on how well the fit indices identified the latent classes. It is more difficult for fit indices to detect class proportions splits with rarer or smaller class sizes. However, the aBIC performed relatively well in these situations across all modelling settings. Model complexity may also factor into the accuracy rate. It was not overly prominent in our study because population B was only slightly more complex than population A. Even so, Figure 4.2 and Table 4.2 reflect slight improvements in the accuracy rates of population A over B. It is difficult to generalize this for more complex models without further investigation.

Based on our findings, covariates generally increased the overall enumeration accuracy (hindered in few cases) and proved to be beneficial for certain modelling situations such as poor quality conditions or class split. However, these improvements can only be guaranteed given that the true covariate specification is used for class enumeration. Realistically, the true specification of the population model is usually unknown and randomly guessing a specification will generally lead to a misspecfied and overestimated LCA model (Nylund-Gibson and Masyn, 2016). Our study indicates that the unconditional model performed equivalently as well as the conditional model with the correct covariate specification. Even with the presence of large covariate effect sizes in the population model, the unconditional model performed as well as the conditional model as illustrated in Figures 4.2 and 4.1. Therefore, it is reasonable and our recommendation to use the unconditional model for class enumeration.

Findings in this study were limited to binary, categorical latent class models and only a limited number of population models with three classes. Future studies could expand these parameters and to other mixture models to understand how model complexity impacts class enumeration and the overall performance of the fit indices. Additionally we only examined a limited number of covariates to one continuous variable and one binary categorical variable. It would be interesting to see how multiple covariates would interplay with class enumeration. Other model section criteria could also be examined such as the AIC3, which in some studies has shown to outperform other information criteria (Fonseca and Cardoso, 2007; Yang and Yang, 2007; Dias, 2006) or alternative cross validation approaches to approximate the LRT distribution given that BLRT performed well. We could also examine under what conditions could low quality items or small samples still be used while ensuring results obtained are still justifiable and unbiased. As another example, a power analysis and Type I error analysis could be conducted on BLRT and aVLMR to compare and better understand their performance over a wider range of conditions.

Appendix A

Mplus Code

These are sample template files used to simulate data for population model B. In `Mplus`, the item-response probabilities are represented by thresholds, $\tau_{r_{mk}}$, and calculated by (2.2.6). The class proportion probabilities are expressed as intercepts of a multinomial logistic model given by (2.2.3).

```
[init]]
    iterators = sample covEffects;
    sample = 200 500 1000 2000;
    covEffects = 0.40 0.90 1.50;
    filename= "MC_Data_Pop2_Split1_HQ.inp";
    outputDirectory= ~/PopTwo/Split1/HighQuality/
                    [[covEffects]]/[[sample]]/Data;
[[/init]]

TITLE: Simulated data -- Population B -- High Quality conditions
MONTECARLO:
     names are u1-u10 x w;
     generate= u1-u10(1);
     categorical= u1-u10;
     genclasses= c(3);
     cutpoints=w(0); !binary covariate
     classes= c(3);
     nobs=[[sample]];
     nrep=500;
     repsave=all;
     save= ~/PopTwo/Split1/HighQuality/
            [[covEffects]]/[[sample]]/Data/[[sample]]*.dat;

MODEL POPULATION:
%OVERALL%
    [x@0]; x@1;
    [w@0]; w@1;
    [c#1@2.70]; [c#2@2.16];
    c#1 ON x*0.90; c#2 ON x*0.90;
    u1 ON w*[[covEffects]]; u2 ON w*[[covEffects]];
```

```
%c#1%
    [u1$1*-2.2 u2$1*-2.2  u3$1*-2.2  u4$1*0 u5$1*2.2
    u6$1*0 u7$1*2.2 u8$1*2.2 u9$1*0.62 u10$1*0.62];
%c#2%
    [u1$1*2.2  u2$1*1.39 u3$1*1.1 u4$1*-2.2 u5$1*-2.2
    u6$1*-2.2 u7$1*1.1 u8$1*1.39 u9$1*1.39 u10$1*1.1];
%c#3%
    [u1$1*1.73  u2$1*1.73  u3$1*0.85  u4$1*0.85  u5$1*1.39
    u6$1*1.1 u7$1*-2.2 u8$1*-2.2 u9$1*-2.2 u10$1*1.39];

MODEL:
%OVERALL%
    [c#1@2.70]; [c#2@2.16];
    c#1 ON x*0.90; c#2 ON x*0.90;
    u1 ON w*[[covEffects]]; u2 ON w*[[covEffects]];

%c#1%
    [u1$1*-2.2  u2$1*-2.2  u3$1*-2.2  u4$1*0 u5$1*2.2
    u6$1*0 u7$1*2.2 u8$1*2.2 u9$1*0.62 u10$1*0.62];
%c#2%
    [u1$1*2.2  u2$1*1.39 u3$1*1.1  u4$1*-2.2 u5$1*-2.2
    u6$1*-2.2 u7$1*1.1  u8$1*1.39 u9$1*1.39 u10$1*1.1];
%c#3%
    [u1$1*1.73  u2$1*1.73  u3$1*0.85  u4$1*0.85  u5$1*1.39
    u6$1*1.1 u7$1*-2.2 u8$1*-2.2 u9$1*-2.2 u10$1*1.39];

ANALYSIS:
    type=mixture;
```

```
[[init]]
    iterators = sample covEffects;
    sample = 200 500 1000 2000;
    covEffects = 0.40 0.90 1.50;
    filename= "MC_Data_Pop2_Split1_MQ.inp";
    outputDirectory= ~/PopTwo/Split1/ModerateQuality/
                      [[covEffects]]/[[sample]]/Data;
[[/init]]

TITLE: Simulated data -- Population B -- Moderate Quality conditions

MONTECARLO:
    names are u1-u10 w x;
    generate= u1-u10(1);
    categorical= u1-u10;
    genclasses= c(3);
    cutpoints=w(0); !binary covariate
    classes= c(3);
    nobs=[[sample]];
    nrep=500;
    repsave=all;
    save= ~/PopTwo/Split2/ModerateQuality/[[covEffects]]/
                        [[sample]]/Data/[[sample]]*.dat;

MODEL POPULATION:
%OVERALL%
    [x@0]; x@1; [w@0]; w@1;
    [c#1@2.70]; [c#2@2.16];
    c#1 ON x*0.90; c#2 ON x*0.90; u1 ON w*[[covEffects]]; u2 ON w*[[covEffects]];
```

```
%c#1%
    [u1$1*-1.1  u2$1*-1.1  u3$1*-1.1  u4$1*0     u5$1*2.2
     u6$1*0     u7$1*2.2   u8$1*2.2   u9$1*0.62  u10$1*0.62];

%c#2%
    [u1$1*0.85  u2$1*1.73  u3$1*0.62  u4$1*-1.39 u5$1*-1.39
     u6$1*-1.39 u7$1*0.85  u8$1*0.41  u9$1*1.39  u10$1*0.41];

%c#3%
    [ u1$1*1.73 u2$1*1.73 u3$1*0.85 u4$1*0.85 u5$1*1.39
    u6$1*1.1  u7$1*-1.1 u8$1*-1.1 u9$1*-1.1 u10$1*0.2];

MODEL
%OVERALL%
    [c#1@2.70]; [c#2@2.16];
    c#1 ON x*0.90; c#2 ON x*0.90;
    u1 ON w*[[covEffects]]; u2 ON w*[[covEffects]];

%c#1%
    [u1$1*-1.1  u2$1*-1.1  u3$1*-1.1  u4$1*0     u5$1*2.2
    u6$1*0     u7$1*2.2   u8$1*2.2   u9$1*0.62  u10$1*0.62];
%c#2%
    [u1$1*0.85  u2$1*1.73  u3$1*0.62  u4$1*-1.39 u5$1*-1.39
    u6$1*-1.39 u7$1*0.85  u8$1*0.41  u9$1*1.39  u10$1*0.41];
%c#3%
    [ u1$1*1.73 u2$1*1.73 u3$1*0.85 u4$1*0.85 u5$1*1.39
    u6$1*1.1  u7$1*-1.1 u8$1*-1.1 u9$1*-1.1 u10$1*0.2];

ANALYSIS:
    type=mixture;
```

```
[[init]]
    iterators = sample covEffects;
    sample = 200 500 1000 2000;
    covEffects = 0.40 0.90 1.50;
    filename= "MC_Data_Pop1_Split1_LQ.inp";
    outputDirectory= ~/PopOne/Split1/LowQuality/
                    [[covEffects]]/[[sample]]/Data;
[[/init]]

TITLE: Simulated data -- Population B -- Low Quality Conditions

MONTECARLO:
     names are u1-u10 w x;
     generate= u1-u10(1);
     categorical= u1-u10;
     genclasses= c(3);
     cutpoints=w(0); !binary covariate
     classes= c(3);
     nobs=[[sample]];
     nrep=500;
     repsave=all;
     save= ~/PopOne/Split1/LowQuality/
            [[covEffects]]/[[sample]]/Data/[[sample]]*.dat;

MODEL POPULATION:
%OVERALL%
    [x@0]; x@1;
    [w@0]; w@1;
    [c#1@2.70]; [c#2@2.16];
    c#1 ON x*0.90; c#2 ON x*0.90;
    u1 ON w*[[covEffects]]; u2 ON w*[[covEffects]];
```

```
%c#1%
    [u1$1*0     u2$1*0     u3$1*0     u4$1*0.85  u5$1*0.41
    u6$1*0     u7$1*0.62  u8$1*0.2   u9$1*1.39  u10$1*0.85];
%c#2%
    [u1$1*0.41  u2$1*0.41  u3$1*0.85  u4$1*0.2   u5$1*0.2
    u6$1*0.2   u7$1*0.85  u8$1*0.41  u9$1*0.85  u10$1*0.41];
%c#3%
    [u1$1*1.73  u2$1*1.73  u3$1*0.85  u4$1*0.85  u5$1*0.62
    u6$1*1.1   u7$1*0     u8$1*0     u9$1*0     u10$1*0.85];

MODEL:
%OVERALL%
    [c#1@2.70]; [c#2@2.16];
    c#1 ON x*0.90; c#2 ON x*0.90;
    u1 ON w*[[covEffects]]; u2 ON w*[[covEffects]];

%c#1%
    [u1$1*0     u2$1*0     u3$1*0     u4$1*0.85  u5$1*0.41
    u6$1*0     u7$1*0.62  u8$1*0.2   u9$1*1.39  u10$1*0.85];
%c#2%
    [u1$1*0.41  u2$1*0.41  u3$1*0.85  u4$1*0.2   u5$1*0.2
    u6$1*0.2   u7$1*0.85  u8$1*0.41  u9$1*0.85  u10$1*0.41];
%c#3%
    [u1$1*1.73  u2$1*1.73  u3$1*0.85  u4$1*0.85  u5$1*0.62
    u6$1*1.1   u7$1*0     u8$1*0     u9$1*0     u10$1*0.85];

ANALYSIS:
    type=mixture;
```

Appendix B

Simulation Results

Table B.1: Proportion of times a non-significant p-value selected a given class model for the aVLMR & BLRT for Population A Split 1. The shaded columns represent the true number of latent classes in the population.

			aVLMR										BLRT									
			No Covariates					Covariates					No Covariates					Covariates				
Quality	Effect	Sample	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
HQ	0.4	200	1	36	56	8	0	1	40	55	5	1	0	23	53	24	0	0	19	54	25	1
		500	0	7	87	7	0	0	8	87	5	1	0	16	82	2	0	0	14	82	3	1
		1000	0	1	90	9	1	0	1	93	7	0	0	1	98	1	0	0	2	95	2	1
		2000	0	0	92	9	0	0	0	93	7	0	0	4	95	1	0	0	0	96	2	1
	0.9	200	0	21	71	7	2	1	23	70	6	1	0	30	65	5	0	0	26	67	6	1
		500	0	2	87	10	2	0	2	91	8	0	0	17	83	0	0	0	10	86	3	1
		1000	0	0	91	9	1	0	0	93	8	0	0	5	94	1	0	0	1	95	2	1
		2000	0	0	89	10	1	0	0	92	8	1	1	6	93	0	0	0	3	94	2	1
	1.5	200	1	8	81	11	0	1	6	85	8	1	0	25	66	9	0	0	30	69	0	1
		500	0	0	88	11	1	0	0	94	7	0	1	0	81	0	18	0	1	82	1	16
		1000	0	0	91	9	1	0	0	92	7	1	0	6	94	0	0	0	2	95	1	1
		2000	0	0	92	8	1	0	0	94	6	1	1	0	94	5	0	0	1	96	1	2
MQ	0.4	200	0	2	83	15	0	1	89	9	2	0	5	42	53	0	0	0	43	55	1	1
		500	0	58	30	2	1	0	60	39	1	1	0	2	91	8	0	0	3	92	4	1
		1000	0	19	75	6	0	0	20	76	4	0	0	0	93	7	0	0	1	95	3	1
		2000	0	0	95	5	0	0	0	93	7	1	0	0	93	7	0	0	1	95	3	1
	0.9	200	2	78	20	1	0	3	80	17	1	1	0	38	62	0	0	0	34	63	1	1
		500	0	43	55	3	0	0	45	54	2	0	0	2	91	8	0	0	3	86	9	1
		1000	0	3	91	7	0	0	45	54	2	0	0	2	93	5	0	0	0	92	6	2
		2000	0	1	89	11	0	0	0	92	8	1	0	2	95	3	0	0	1	96	2	1
	1.5	200	5	66	29	2	0	3	52	43	2	0	0	36	64	0	0	0	37	60	1	1
		500	1	16	77	7	0	0	4	92	4	0	0	2	88	9	1	0	3	89	6	2
		1000	0	0	93	8	0	0	0	91	9	1	0	1	89	10	0	0	2	90	6	1
		2000	0	0	90	10	1	0	0	91	9	0	0	1	90	9	0	0	2	91	5	1
LQ	0.4	200	0	90	9	2	0	92	8	0	0	0	89	8	0	3	0	84	9	1	4	1
		500	95	6	0	0	0	92	9	0	0	0	81	17	1	1	0	76	18	2	2	1
		1000	94	6	1	0	0	92	8	1	0	0	79	13	8	0	0	67	20	9	3	1
		2000	89	11	1	0	0	93	7	1	0	0	73	20	7	0	0	67	21	9	2	1
	0.9	200	91	8	2	0	0	91	10	0	0	0	88	10	2	0	0	77	17	3	1	1
		500	95	6	0	0	0	92	8	0	0	0	76	19	5	0	0	72	20	6	2	0
		1000	95	5	0	0	0	91	8	2	0	0	49	42	9	0	0	51	39	8	1	1
		2000	86	14	1	0	0	87	13	1	0	0	48	43	9	0	0	46	40	11	2	1
	1.5	200	90	10	1	0	0	91	8	1	1	0	79	19	2	0	0	69	23	3	4	1
		500	94	7	0	0	0	88	11	1	0	0	70	30	0	0	0	64	34	1	1	0
		1000	89	11	1	0	0	64	35	1	0	0	31	65	4	0	0	23	68	7	1	1
		2000	79	20	2	0	0	20	76	3	1	0	30	62	7	1	0	27	61	8	2	1

Table B.2: Proportion of times a non-significant p-value selected a given class model for the aVLMR & BLRT for Population A Split 2. The shaded columns represent the true number of latent classes in the population.

			aVLMR										BLRT									
			No Covariates					Covariates					No Covariates					Covariates				
Quality	Effect	Sample	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
HQ	0.4	200	0	0	80	16	2	1	1	84	13	1	0	23	53	24	0	0	19	54	25	1
		500	0	0	88	11	0	0	0	89	11	0	0	16	82	2	0	0	14	82	3	1
		1000	0	0	91	8	1	0	0	94	6	0	0	1	98	1	0	0	2	95	2	1
		2000	0	0	86	13	0	0	0	93	7	0	0	4	95	1	0	0	0	96	2	1
	0.9	200	0	1	83	15	1	0	0	84	15	1	0	30	65	5	0	0	26	67	6	1
		500	0	0	88	10	2	0	0	94	6	0	0	17	83	0	0	0	10	86	3	1
		1000	0	0	90	10	0	0	0	93	7	0	0	5	94	1	0	0	1	95	2	1
		2000	0	0	89	11	0	0	0	90	10	0	1	6	93	0	0	0	3	94	2	1
	1.5	200	0	0	78	21	1	1	0	81	16	2	0	25	66	9	0	0	30	69	0	1
		500	0	0	85	13	2	0	0	90	10	0	1	0	81	0	18	0	1	82	1	16
		1000	0	0	84	14	2	0	0	89	11	0	0	6	94	0	0	0	2	95	1	1
		2000	0	0	88	9	3	0	0	91	9	0	1	0	94	5	0	0	1	96	1	2
MQ	0.4	200	3	29	63	5	0	3	37	56	4	0	5	42	53	0	0	0	43	55	1	1
		500	0	0	92	8	0	0	0	95	4	1	0	2	91	8	0	0	3	92	4	1
		1000	0	0	90	10	0	0	0	94	6	0	0	0	93	7	0	0	1	95	3	1
		2000	0	0	92	6	2	0	0	92	7	1	0	0	93	7	0	0	1	95	3	1
	0.9	200	8	16	73	3	0	9	15	70	6	0	0	38	62	0	0	0	34	63	1	1
		500	0	1	91	8	0	1	89	0	8	1	0	2	91	8	0	0	3	86	9	1
		1000	0	0	86	13	1	0	0	91	8	1	0	2	93	5	0	0	0	92	6	2
		2000	0	0	89	10	1	0	0	95	5	0	0	2	95	3	0	0	1	96	2	1
	1.5	200	8	15	67	8	2	10	3	80	7	0	0	36	64	0	0	0	37	60	1	1
		500	0	0	94	6	0	4	93	3	0	0	0	2	88	9	1	0	3	89	6	2
		1000	0	0	94	4	2	0	0	92	7	1	0	1	89	10	0	0	2	90	6	1
		2000	0	0	94	6	0	1	0	94	4	1	0	1	90	9	0	0	2	91	5	1
LQ	0.4	200	93	6	1	0	0	91	9	0	0	0	89	8	0	3	0	84	9	1	4	1
		500	88	10	2	0	0	92	8	0	0	0	81	17	1	1	0	76	18	2	2	1
		1000	77	21	1	1	0	80	20	0	0	0	79	13	8	0	0	67	20	9	3	1
		2000	40	57	3	0	0	35	60	4	1	0	73	20	7	0	0	67	21	9	2	1
	0.9	200	92	8	0	0	0	91	9	0	0	0	88	10	2	0	0	77	17	3	1	1
		500	92	8	0	0	0	77	21	2	0	0	76	19	5	0	0	72	20	6	2	0
		1000	61	38	1	0	0	35	61	4	0	0	49	42	9	0	0	51	39	8	1	1
		2000	27	66	7	0	0	2	87	11	0	0	48	43	9	0	0	46	40	11	2	1
	1.5	200	91	9	0	0	0	83	15	2	0	0	79	19	2	0	0	69	23	3	4	1
		500	87	13	0	0	0	31	65	4	0	0	70	30	0	0	0	64	34	1	1	0
		1000	56	42	1	0	1	3	88	9	0	0	31	65	4	0	0	23	68	7	1	1
		2000	8	83	8	1	0	0	0	94	5	1	30	62	7	1	0	27	61	8	2	1

Table B.3: Proportion of times a non-significant p-value selected a given class model for the aVLMR & BLRT for Population B Split 1. The shaded columns represent the true number of latent classes in the population.

			aVLMR										BLRT									
			No Covariates					Covariates					No Covariates					Covariates				
Quality	Effect	Sample	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
HQ	0.4	200	1	21	66	11	1	1	22	67	10	0	0	38	58	0	4	1	40	59	0	0
		500	0	2	88	10	0	0	2	29	0	69	0	22	68	0	10	0	22	68	0	10
		1000	0	0	88	11	0	0	0	95	5	0	0	0	93	0	7	0	0	95	0	5
		2000	0	0	91	10	1	0	0	94	6	0	0	1	93	0	6	0	1	96	0	3
	0.9	200	1	23	66	10	1	1	22	70	7	0	0	10	60	30	0	0	17	53	30	0
		500	0	1	87	12	1	0	2	89	8	1	0	23	77	0	0	0	20	80	0	0
		1000	0	0	90	9	1	0	0	84	16	0	0	8	92	0	0	3	3	94	0	0
		2000	0	0	89	10	1	0	0	84	5	1	0	9	91	0	0	0	9	91	0	0
	1.5	200	1	24	67	7	1	3	27	66	4	0	0	11	89	0	0	0	11	89	0	0
		500	0	2	86	11	1	0	3	89	8	0	0	0	93	7	0	0	0	95	5	0
		1000	0	11	88	1	0	0	0	88	11	1	0	0	96	4	0	0	0	100	0	0
		2000	0	0	86	13	1	0	0	97	3	0	0	0	97	3	0	0	0	100	0	0
MQ	0.4	200	2	80	19	1	0	0	76	20	4	0	0	34	60	6	0	0	34	61	5	0
		500	0	44	52	4	0	0	32	68	0	0	0	0	91	6	3	0	0	92	5	3
		1000	0	4	91	5	1	0	0	93	7	0	0	0	89	11	0	0	0	90	10	0
		2000	0	0	91	8	1	0	0	92	8	0	0	0	89	11	0	0	0	90	10	0
	0.9	200	2	78	20	1	0	0	84	16	0	0	0	34	66	0	0	0	32	67	1	0
		500	0	45	53	2	0	0	40	56	4	0	0	0	91	9	0	0	0	91	9	0
		1000	0	4	91	4	1	0	8	88	4	0	0	0	86	14	0	0	0	90	10	0
		2000	0	0	92	8	0	0	0	100	0	0	0	0	86	14	0	0	0	87	13	0
	1.5	200	2	81	17	1	0	0	88	8	4	0	0	0	34	66	0	0	0	70	30	0
		500	0	48	49	4	0	0	32	68	0	0	0	0	94	6	0	0	0	95	5	0
		1000	0	5	89	6	0	0	4	92	4	0	0	0	90	10	0	0	0	92	6	2
		2000	0	0	85	15	1	0	0	96	4	0	0	1	91	8	0	0	0	93	7	0
LQ	0.4	200	92	8	1	0	0	94	6	0	0	0	91	8	0	1	0	95	5	0	1	0
		500	91	8	0	1	0	92	7	1	0	0	80	19	1	0	0	82	16	2	0	0
		1000	94	6	0	0	0	92	8	0	0	0	79	14	7	0	0	75	25	0	0	0
		2000	83	16	1	0	0	81	17	2	0	0	71	21	8	0	0	71	20	9	0	0
	0.9	200	91	8	1	0	0	90	9	1	0	0	89	11	0	0	0	90	9	1	0	0
		500	93	7	0	0	0	88	12	0	0	0	77	20	3	0	0	77	20	3	0	0
		1000	89	11	0	0	0	89	11	0	0	0	51	43	6	0	0	60	34	6	0	0
		2000	78	27	1	0	0	79	21	0	0	0	49	44	7	0	0	49	44	7	0	0
	1.5	200	88	12	1	0	0	90	8	2	0	0	80	20	0	0	0	79	20	0	1	0
		500	82	18	0	0	0	90	10	0	0	0	58	34	8	0	0	51	44	5	0	0
		1000	92	36	2	0	0	86	10	4	0	0	25	66	10	0	0	29	59	12	0	0
		2000	23	73	4	0	0	78	22	0	0	0	28	62	10	0	0	30	57	13	0	0

Table B.4: Proportion of times a non-significant p-value selected a given class model for the aVLMR & BLRT for Population B Split 2. The shaded columns represent the true number of latent classes in the population.

			aVLMR										BLRT									
			No Covariates					Covariates					No Covariates					Covariates				
Quality	Effect	Sample	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
HQ	0.4	200	1	0	82	14	3	0	89	9	0	1	0	0	89	11	0	0	0	86	14	0
		500	0	0	87	12	1	0	0	93	6	1	0	0	91	6	3	0	0	93	7	0
		1000	0	0	87	12	1	0	0	92	8	1	0	0	95	5	0	0	0	95	5	0
		2000	0	0	89	10	1	0	0	89	10	1	0	0	97	3	0	0	0	100	5	0
	0.9	200	0	0	80	16	3	1	0	88	10	1	0	0	89	11	0	0	0	95	5	0
		500	0	0	88	12	1	0	0	93	6	1	0	0	89	9	3	0	0	93	7	0
		1000	0	0	88	12	1	0	0	91	8	1	0	0	91	9	0	0	0	86	14	0
		2000	0	0	87	12	0	0	0	91	8	1	0	0	93	7	0	0	0	100	0	0
	1.5	200	1	0	82	15	1	1	0	87	12	1	0	0	89	11	0	0	0	93	7	0
		500	0	0	94	14	2	0	0	92	7	0	0	0	86	14	0	0	0	86	14	0
		1000	0	0	86	12	2	0	0	92	7	1	0	0	91	9	0	0	0	100	0	0
		2000	0	0	82	17	1	0	0	92	7	1	0	0	92	8	0	0	0	100	0	0
MQ	0.4	200	6	22	65	7	0	8	21	67	4	0	0	0	80	20	0	0	0	97	3	0
		500	0	0	93	7	0	0	0	95	5	0	0	0	86	14	0	0	0	94	6	0
		1000	0	0	93	7	0	0	0	93	6	1	0	0	94	6	0	0	0	97	3	0
		2000	0	0	92	8	0	0	0	94	5	1	0	0	90	10	0	0	0	97	3	0
	0.9	200	6	23	65	6	0	9	22	66	3	0	0	0	78	12	0	0	0	97	3	0
		500	0	0	92	8	1	0	0	95	5	0	0	0	86	14	0	0	0	94	6	0
		1000	0	0	92	8	0	0	0	94	5	0	0	0	89	6	6	0	0	94	6	0
		2000	0	0	88	12	0	0	0	94	6	0	0	0	88	12	0	0	0	100	0	0
	1.5	200	10	25	59	5	0	10	22	64	4	0	0	0	100	0	0	0	0	94	6	0
		500	0	0	91	8	1	1	0	92	6	0	0	0	83	17	0	0	0	94	6	0
		1000	0	0	92	8	0	0	0	95	5	0	0	0	86	14	0	0	0	94	6	0
		2000	0	0	84	16	1	0	0	96	4	0	0	0	87	13	0	0	0	94	6	0
LQ	0.4	200	90	10	0	0	0	90	10	1	0	0	83	17	0	0	0	73	23	3	0	0
		500	89	10	0	0	0	80	19	1	0	0	60	40	0	0	0	27	70	3	0	0
		1000	67	31	1	0	0	32	62	5	0	0	20	77	3	0	0	3	93	3	0	0
		2000	19	76	5	0	0	1	91	8	0	0	0	64	36	0	0	0	95	5	0	0
	0.9	200	88	11	1	0	0	89	10	1	0	0	66	34	0	0	0	63	37	0	0	0
		500	82	17	1	0	0	79	20	2	0	0	43	54	3	0	0	13	83	3	0	0
		1000	54	44	2	0	0	27	68	4	1	0	3	91	6	0	0	0	83	13	3	0
		2000	8	88	3	0	0	1	90	8	0	0	1	92	7	0	0	0	86	14	0	0
	1.5	200	87	12	1	0	0	91	9	0	0	0	57	40	3	0	0	64	36	0	0	0
		500	71	29	0	0	0	80	19	1	0	0	14	80	6	0	0	14	86	0	0	0
		1000	26	72	3	0	0	26	70	5	0	0	3	91	6	0	0	0	83	17	0	0
		2000	4	90	6	0	0	1	91	8	1	0	0	94	6	0	0	0	84	16	0	0

Table B.5: Proportion of times a given model is selected based on the lowest ICL and CLC value for Population A Split 1. The shaded columns represent the true number of latent classes in the population.

			ICL								CLC							
			No Covariates				Covariates				No Covariates				Covariates			
Quality	Effect	Sample	2	3	4	5	2	3	4	5	2	3	4	5	2	3	4	5
HQ	0.4	200	41	32	21	7	49	28	18	6	0	1	11	89	0	1	9	91
		500	1	16	37	48	0	21	32	48	0	0	13	88	0	0	9	92
		1000	0	10	37	54	0	10	30	61	0	0	14	87	0	0	6	95
		2000	0	3	37	61	0	4	24	73	0	0	19	81	0	0	9	92
	0.9	200	23	46	26	6	25	48	21	7	0	1	6	94	0	0	5	95
		500	0	14	38	49	0	18	37	46	0	0	11	89	0	0	8	92
		1000	0	9	34	58	0	13	30	58	0	0	13	87	0	0	9	92
		2000	0	5	32	64	0	6	30	64	0	0	12	88	0	0	13	88
	1.5	200	3	71	22	5	1	74	20	5	0	0	7	93	0	0	3	98
		500	0	24	33	43	0	22	30	49	0	0	8	92	0	0	8	93
		1000	0	12	30	59	0	17	23	61	0	0	9	92	0	0	7	93
		2000	0	5	18	78	0	6	22	73	0	0	11	89	0	0	10	91
MQ	0.4	200	83	11	6	1	89	9	2	0	0	1	17	82	0	1	12	88
		500	19	18	33	31	23	13	29	36	0	2	16	83	0	1	16	84
		1000	1	12	37	51	1	13	34	53	0	1	18	82	0	0	17	84
		2000	0	3	27	71	0	5	24	72	0	0	15	86	0	0	12	88
	0.9	200	82	12	7	0	85	11	5	0	0	0	15	86	0	0	8	92
		500	11	21	37	33	9	22	30	39	0	0	13	87	0	1	10	90
		1000	0	12	31	58	0	19	28	54	0	1	17	82	0	0	10	90
		2000	0	4	26	70	0	8	28	64	0	0	15	86	0	0	14	87
	1.5	200	67	26	7	1	54	38	7	1	0	1	14	86	0	0	7	93
		500	1	37	34	28	0	27	40	34	0	1	15	85	0	0	16	85
		1000	0	17	30	54	0	19	30	51	0	1	15	85	0	0	11	89
		2000	0	2	24	75	0	4	21	76	0	0	10	90	0	0	8	92
LQ	0.4	200	42	37	18	3	60	27	12	2	1	5	25	70	3	4	20	74
		500	12	25	30	34	27	24	24	27	2	8	22	69	6	9	20	67
		1000	7	13	27	53	12	20	31	38	1	7	21	72	3	13	27	58
		2000	7	14	25	55	9	20	28	43	3	11	22	65	5	17	27	52
	0.9	200	41	36	19	6	57	26	15	2	0	4	20	76	0	4	18	79
		500	15	24	31	31	22	25	28	26	2	7	24	68	3	9	24	65
		1000	6	22	29	45	11	22	27	41	3	12	24	63	4	13	24	60
		2000	6	7	32	56	11	19	28	43	4	6	30	61	6	15	26	54
	1.5	200	45	33	20	3	59	26	13	4	1	4	23	73	2	5	25	69
		500	12	25	30	34	21	26	22	32	1	10	20	70	3	8	19	71
		1000	6	13	33	49	8	15	34	44	3	5	27	66	3	6	29	63
		2000	3	14	29	55	4	15	24	59	2	9	26	64	2	8	20	71

Table B.6: Proportion of times a given model is selected based on the lowest ICL and CLC value for Population A Split 2. The shaded columns represent the true number of latent classes in the population.

			ICL								CLC							
			No Covariates				Covariates				No Covariates				Covariates			
Quality	Effect	Sample	2	3	4	5	2	3	4	5	2	3	4	5	2	3	4	5
HQ	0.4	200	0	86	11	3	0	89	8	3	0	0	2	98	0	0	0	100
		500	0	33	25	42	0	32	25	43	0	0	7	93	0	0	2	98
		1000	0	11	26	63	0	12	23	65	0	0	10	90	0	0	4	96
		2000	0	5	23	72	0	4	26	70	0	0	11	89	0	0	9	91
	0.9	200	0	88	12	0	0	92	7	1	0	0	2	98	0	0	1	99
		500	0	33	34	33	0	44	28	28	0	0	8	92	0	0	5	95
		1000	0	8	27	65	0	10	31	59	0	0	5	95	0	0	5	95
		2000	0	1	25	74	0	5	17	78	0	0	11	89	0	0	5	95
	1.5	200	0	96	4	0	0	94	6	0	0	0	0	100	0	0	1	99
		500	0	33	31	36	0	39	28	33	0	0	6	94	0	0	6	94
		1000	0	10	29	61	0	21	26	53	0	0	7	93	0	0	2	98
		2000	0	4	20	76	0	7	23	70	0	0	6	94	0	0	7	93
MQ	0.4	200	12	76	11	1	10	87	3	0	0	1	15	84	0	0	7	93
		500	0	30	39	31	0	37	28	35	0	0	18	82	0	0	10	90
		1000	0	11	35	54	0	18	34	48	0	0	14	86	0	0	8	92
		2000	0	2	25	73	0	9	21	70	0	0	11	89	0	0	10	90
	0.9	200	1	94	4	1	0	96	4	0	0	1	16	83	0	1	12	87
		500	0	44	31	25	0	47	29	24	0	1	9	90	0	0	5	95
		1000	0	15	31	54	0	18	29	53	0	0	12	88	0	0	8	92
		2000	0	3	21	76	0	8	23	69	0	1	8	91	0	0	9	91
	1.5	200	0	96	4	0	1	97	2	0	0	0	15	85	0	0	4	96
		500	0	42	31	27	0	47	24	29	0	0	14	86	0	1	7	92
		1000	0	12	36	52	0	13	28	59	0	1	13	86	0	0	5	95
		2000	0	3	28	69	0	7	20	73	0	0	15	85	0	0	6	94
LQ	0.4	200	48	30	22	0	59	22	14	5	2	7	24	67	0	4	22	74
		500	19	20	29	32	22	29	24	25	1	6	23	70	2	15	20	63
		1000	13	18	19	50	14	21	29	36	3	10	16	71	5	13	31	51
		2000	3	12	20	65	9	15	29	47	1	7	17	75	5	13	24	58
	0.9	200	41	38	18	3	54	32	11	3	1	2	15	82	2	11	18	69
		500	15	22	33	30	24	24	30	22	0	13	21	66	1	9	32	58
		1000	13	14	28	45	14	28	31	27	4	5	29	62	6	11	29	54
		2000	6	10	26	58	2	13	32	53	1	7	26	66	1	8	27	64
	1.5	200	42	41	12	5	73	19	6	2	0	4	23	73	0	7	21	72
		500	17	26	25	32	27	30	24	19	3	9	27	61	0	12	26	62
		1000	9	18	34	39	9	15	35	41	3	9	31	57	2	6	22	70
		2000	2	13	23	62	1	15	24	60	0	5	20	75	0	8	24	6

Table B.7: Proportion of times a given model is selected based on the lowest ICL and CLC value for Population B Split 1. The shaded columns represent the true number of latent classes in the population.

			ICL								CLC							
			No Covariates				Covariates				No Covariates				Covariates			
Quality	Effect	Sample	2	3	4	5	2	3	4	5	2	3	4	5	2	3	4	5
HQ	0.4	200	22	45	23	11	24	59	14	3	0	0	7	94	0	0	3	97
		500	0	20	32	48	0	29	33	38	0	0	10	90	0	0	13	87
		1000	0	9	31	61	0	29	33	38	0	0	12	89	0	0	16	87
		2000	0	2	27	71	0	7	32	61	0	0	13	87	0	0	13	87
	0.9	200	24	43	22	11	27	51	15	7	0	0	5	95	0	0	7	93
		500	0	23	34	43	1	29	31	39	0	0	11	89	0	0	7	93
		1000	0	8	28	64	0	7	36	57	0	0	10	90	0	0	14	86
		2000	0	3	30	68	0	7	26	67	0	0	14	86	0	0	13	87
	1.5	200	29	42	25	6	33	51	12	4	0	0	9	92	0	0	4	96
		500	0	24	37	40	0	23	37	40	0	0	11	89	0	0	7	93
		1000	0	9	33	58	0	15	32	53	0	0	14	86	0	0	8	92
		2000	0	2	29	69	0	4	24	72	0	0	13	87	0	0	11	89
MQ	0.4	200	80	13	8	0	80	12	8	0	0	1	11	88	0	4	8	88
		500	10	24	31	36	12	40	20	28	0	0	16	84	0	0	16	84
		1000	0	11	36	53	0	20	4	76	0	0	17	83	0	0	4	96
		2000	0	4	24	72	0	8	28	64	0	0	14	86	0	0	20	80
	0.9	200	83	11	6	0	88	0	12	0	0	1	13	86	0	0	8	92
		500	11	25	29	35	20	36	4	40	0	1	15	84	0	0	4	96
		1000	0	11	34	55	0	12	20	68	0	1	17	83	0	0	4	96
		2000	0	5	24	72	0	4	20	76	0	0	13	87	0	0	8	92
	1.5	200	84	10	7	0	88	8	4	0	0	2	13	86	0	0	8	92
		500	15	24	31	30	16	20	40	24	0	1	16	83	0	0	12	88
		1000	0	13	30	57	0	16	32	52	0	1	15	85	0	0	20	80
		2000	0	3	24	73	0	12	8	80	0	0	13	87	0	0	8	92
LQ	0.4	200	45	34	16	6	24	20	32	24	1	6	21	72	1	4	18	77
		500	15	22	32	32	24	20	32	24	2	5	24	70	5	4	28	63
		1000	11	15	26	48	14	15	28	43	4	7	24	66	5	8	22	65
		2000	7	16	25	53	10	16	34	40	4	13	23	61	4	14	35	46
	0.9	200	42	32	19	7	52	23	20	5	1	6	25	69	2	6	16	76
		500	14	22	27	38	14	21	30	35	3	7	23	68	5	18	28	49
		1000	9	15	27	49	14	21	30	35	4	9	23	65	5	18	28	49
		2000	7	14	25	54	7	22	31	40	4	11	23	63	1	15	32	52
	1.5	200	46	35	16	3	56	22	14	8	1	6	23	70	2	4	14	80
		500	16	22	29	33	28	26	24	22	2	11	23	64	2	10	24	64
		1000	8	16	33	43	12	22	26	40	3	9	28	61	2	14	22	62
		2000	6	13	25	56	2	24	22	52	3	10	23	64	0	14	26	60

Table B.8: Proportion of times a given model is selected based on the lowest ICL and CLC value for Population B Split 2. The shaded columns represent the true number of latent classes in the population.

			ICL								CLC							
			No Covariates				Covariates				No Covariates				Covariates			
Quality	Effect	Sample	2	3	4	5	2	3	4	5	2	3	4	5	2	3	4	5
HQ	0.4	200	0	89	10	1	0	93	7	0	0	0	4	96	0	0	3	97
		500	0	33	31	36	0	38	29	33	0	0	6	94	0	0	4	96
		1000	0	10	23	67	0	14	27	59	0	0	5	95	0	0	6	94
		2000	0	4	20	76	0	7	19	74	0	0	8	92	0	0	6	94
	0.9	200	0	90	9	1	0	93	7	0	0	0	2	98	0	0	2	98
		500	0	30	31	38	0	35	32	33	0	0	6	94	0	0	5	95
		1000	0	9	30	60	0	16	26	58	0	0	8	92	0	0	7	93
		2000	0	3	19	78	0	6	22	72	0	0	8	92	0	0	8	92
	1.5	200	0	90	10	1	0	93	7	0	0	0	3	97	0	0	2	98
		500	0	33	31	36	0	37	31	32	0	0	8	92	0	0	5	95
		1000	0	8	29	63	0	14	26	60	0	0	10	90	0	0	4	96
		2000	0	2	18	80	0	7	24	69	0	0	7	93	0	0	9	91
MQ	0.4	200	4	87	8	0	2	93	4	0	0	1	10	90	0	0	8	92
		500	0	43	31	27	0	53	27	20	0	1	14	86	0	0	11	89
		1000	0	13	29	57	0	23	30	47	0	0	11	89	0	0	11	89
		2000	0	3	21	76	0	7	23	70	0	0	10	90	0	0	8	92
	0.9	200	4	89	7	0	3	91	6	0	0	0	12	88	0	0	6	94
		500	0	40	34	26	0	48	27	25	0	0	15	84	0	0	10	90
		1000	0	13	30	57	0	19	29	52	0	0	11	89	0	0	8	91
		2000	0	4	20	77	0	7	21	73	0	0	10	90	0	0	7	93
	1.5	200	4	88	7	0	3	92	5	0	0	1	13	87	0	0	8	92
		500	0	46	33	21	0	51	27	22	0	1	15	84	0	0	10	90
		1000	0	11	34	55	0	23	24	53	0	0	11	89	0	0	9	91
		2000	0	2	22	76	0	8	19	73	0	0	9	91	0	0	5	95
LQ	0.4	200	43	36	17	4	61	27	9	2	1	7	25	67	2	7	18	73
		500	18	21	28	34	27	28	23	22	3	7	21	69	3	12	24	60
		1000	12	19	26	44	12	20	26	42	2	10	22	66	2	9	25	64
		2000	4	13	28	55	3	13	27	56	1	7	24	67	1	7	24	68
	0.9	200	49	32	15	5	60	25	13	2	1	5	23	71	2	7	20	72
		500	19	24	28	28	29	28	22	21	1	9	26	64	4	13	22	61
		1000	12	18	28	43	11	18	32	39	3	9	25	64	2	8	29	61
		2000	2	15	25	58	2	13	29	56	1	9	21	69	0	7	26	67
	1.5	200	50	35	12	3	60	25	13	2	0	5	21	74	2	7	20	72
		500	21	23	27	29	29	28	22	21	3	8	24	65	4	13	22	61
		1000	9	20	27	45	11	18	32	39	2	9	21	68	2	8	29	61
		2000	3	13	28	56	2	13	29	56	1	8	24	67	0	7	26	67

Table B.9: Proportion of times a given model is selected based on the lowest AIC and CAIC value for Population A Split 1. The shaded columns represent the true number of latent classes in the population.

			AIC								CAIC							
			No Covariates				Covariates				No Covariates				Covariates			
Quality	Effect	Sample	2	3	4	5	2	3	4	5	2	3	4	5	2	3	4	5
HQ	0.4	200	1	61	31	8	1	53	35	12	72	29	0	0	77	23	0	0
		500	0	59	29	13	0	53	30	17	7	94	0	0	10	91	0	0
		1000	0	61	24	16	0	51	33	17	0	100	0	0	0	100	0	0
		2000	0	53	28	20	0	49	31	21	0	100	0	0	0	100	0	0
	0.9	200	1	62	29	9	1	52	32	16	47	54	0	0	46	55	0	0
		500	0	54	34	13	0	52	34	15	1	99	0	0	0	100	0	0
		1000	0	55	29	16	0	50	34	17	0	100	0	0	0	100	0	0
		2000	0	53	26	22	0	48	35	18	0	100	0	0	0	100	0	0
	1.5	200	0	59	31	11	0	51	34	15	12	88	0	0	6	94	0	0
		500	0	55	32	14	0	49	33	18	0	100	0	0	0	100	0	0
		1000	0	56	30	15	0	48	35	17	0	100	0	0	0	100	0	0
		2000	0	47	37	16	0	46	37	17	0	100	0	0	0	100	0	0
MQ	0.4	200	15	50	23	14	10	38	30	22	100	0	0	0	100	0	0	0
		500	2	47	34	18	1	42	28	30	96	4	0	0	96	4	0	0
		1000	0	50	30	21	0	41	34	26	57	44	0	0	60	40	0	0
		2000	0	47	33	21	0	35	36	30	2	98	0	0	3	98	0	0
	0.9	200	7	53	28	14	5	42	32	22	100	0	0	0	100	1	0	0
		500	1	49	30	21	0	41	30	30	86	15	0	0	76	24	0	0
		1000	0	48	33	20	0	38	39	24	19	82	0	0	8	93	0	0
		2000	0	47	31	22	0	40	30	31	0	100	0	0	0	100	0	0
	1.5	200	3	50	31	17	0	37	34	30	97	4	0	0	81	20	0	0
		500	0	44	35	22	0	32	37	32	37	63	0	0	9	91	0	0
		1000	0	46	33	22	0	37	31	32	2	99	0	0	0	100	0	0
		2000	0	48	32	21	0	37	35	29	0	100	0	0	0	100	0	0
LQ	0.4	200	76	15	5	4	51	23	14	13	100	0	0	0	100	0	0	0
		500	73	20	3	5	46	20	16	20	100	0	0	0	100	0	0	0
		1000	65	25	9	2	37	25	19	20	100	0	0	0	100	0	0	0
		2000	54	26	15	6	28	23	24	26	100	0	0	0	100	0	0	0
	0.9	200	78	13	5	4	47	24	14	16	100	0	0	0	100	0	0	0
		500	70	21	6	4	43	22	16	20	100	0	0	0	100	0	0	0
		1000	61	27	10	4	33	27	19	22	100	0	0	0	100	0	0	0
		2000	54	25	14	8	20	22	25	34	100	0	0	0	100	0	0	0
	1.5	200	76	12	6	7	44	22	16	19	100	0	0	0	100	0	0	0
		500	71	20	6	4	28	30	23	21	100	0	0	0	100	0	0	0
		1000	56	31	9	5	20	27	24	30	100	0	0	0	100	0	0	0
		2000	43	32	16	11	15	20	31	36	100	0	0	0	100	0	0	0

Table B.10: Proportion of times a given model is selected based on the lowest AIC and CAIC value for Population A Split 2. The shaded columns represent the true number of latent classes in the population.

			AIC								CAIC							
			No Covariates				Covariates				No Covariates				Covariates			
Quality	Effect	Sample	2	3	4	5	2	3	4	5	2	3	4	5	2	3	4	5
HQ	0.4	200	0	51	36	13	0	46	37	17	0	100	0	0	0	100	0	0
		500	0	45	32	23	0	40	36	24	0	100	0	0	0	100	0	0
		1000	0	51	35	14	0	50	29	21	0	100	0	0	0	100	0	0
		2000	0	45	32	23	0	41	30	29	0	100	0	0	0	100	0	0
	0.9	200	0	54	37	9	0	42	44	14	0	100	0	0	0	100	0	0
		500	0	44	37	19	0	29	44	27	0	100	0	0	0	100	0	0
		1000	0	48	35	17	0	40	41	19	0	100	0	0	0	100	0	0
		2000	0	42	38	20	0	43	39	18	0	100	0	0	0	100	0	0
	1.5	200	0	48	43	9	0	40	38	22	0	100	0	0	0	100	0	0
		500	0	45	37	18	0	44	34	22	0	100	0	0	0	100	0	0
		1000	0	44	40	16	0	34	46	20	0	100	0	0	0	100	0	0
		2000	0	54	32	14	0	38	41	21	0	100	0	0	0	100	0	0
MQ	0.4	200	0	49	33	18	0	42	32	26	60	40	0	0	66	34	0	0
		500	0	35	43	22	0	30	44	26	0	100	0	0	0	100	0	0
		1000	0	35	41	24	0	28	42	30	0	100	0	0	0	100	0	0
		2000	0	47	32	21	0	30	41	29	0	100	0	0	0	100	0	0
	0.9	200	0	51	32	17	0	41	36	23	47	53	0	0	36	64	0	0
		500	0	43	28	29	0	29	42	29	1	99	0	0	0	100	0	0
		1000	0	39	39	22	0	28	42	30	0	100	0	0	0	100	0	0
		2000	0	38	34	28	0	25	39	36	0	100	0	0	0	100	0	0
	1.5	200	0	42	36	22	0	32	39	29	32	68	0	0	6	94	0	0
		500	0	45	36	19	0	34	34	32	0	100	0	0	0	100	0	0
		1000	0	39	38	23	0	24	36	40	0	100	0	0	0	100	0	0
		2000	0	40	37	23	0	26	39	35	0	100	0	0	0	100	0	0
LQ	0.4	200	72	18	6	4	32	27	22	19	100	0	0	0	100	0	0	0
		500	51	28	12	9	34	22	19	25	100	0	0	0	100	0	0	0
		1000	39	36	14	11	22	27	24	27	100	0	0	0	100	0	0	0
		2000	32	42	15	11	17	29	23	31	100	0	0	0	100	0	0	0
	0.9	200	71	17	6	6	38	23	16	23	100	0	0	0	100	0	0	0
		500	49	33	13	5	27	33	23	17	100	0	0	0	100	0	0	0
		1000	41	36	11	12	18	34	22	26	100	0	0	0	100	0	0	0
		2000	37	31	13	19	19	31	21	29	100	0	0	0	100	0	0	0
	1.5	200	70	16	9	5	32	20	23	25	100	0	0	0	100	0	0	0
		500	47	34	12	7	28	23	29	20	100	0	0	0	100	0	0	0
		1000	39	37	18	6	15	29	26	30	100	0	0	0	100	0	0	0
		2000	38	28	21	13	12	30	24	34	100	0	0	0	100	0	0	0

Table B.11: Proportion of times a given model is selected based on the lowest AIC and CAIC value for Population B Split 1. The shaded columns represent the true number of latent classes in the population.

			AIC								CAIC							
			No Covariates				Covariates				No Covariates				Covariates			
Quality	Effect	Sample	2	3	4	5	2	3	4	5	2	3	4	5	2	3	4	5
HQ	0.4	200	1	59	30	11	0	57	33	10	49	51	0	0	47	53	0	0
		500	0	57	30	14	0	52	34	14	2	98	0	0	2	98	0	0
		1000	0	49	33	19	0	44	34	22	0	100	0	0	0	100	0	0
		2000	0	54	29	17	0	43	37	20	0	100	0	0	0	100	0	0
	0.9	200	1	58	32	10	0	51	36	13	50	50	0	0	47	53	0	0
		500	0	54	31	15	0	51	29	20	2	98	0	0	2	98	0	0
		1000	0	48	33	20	0	38	43	19	0	100	0	0	0	100	0	0
		2000	0	48	30	22	0	39	33	28	0	100	0	0	0	100	0	0
	1.5	200	1	57	32	11	0	50	36	14	52	48	0	0	48	52	0	0
		500	0	50	34	16	0	43	34	2	2	98	0	0	2	98	0	0
		1000	0	45	33	22	0	50	28	22	0	100	0	0	0	100	0	0
		2000	0	38	31	31	0	42	35	23	0	100	0	0	0	100	0	0
MQ	0.4	200	10	48	27	15	8	40	40	12	100	0	0	0	100	0	0	0
		500	0	48	34	18	0	24	40	36	84	16	0	0	76	24	0	0
		1000	0	49	31	21	0	48	20	32	22	78	0	0	8	92	0	0
		2000	0	47	34	20	0	40	32	28	0	100	0	0	0	100	0	0
	0.9	200	10	45	30	16	12	36	24	28	100	0	0	0	100	0	0	0
		500	0	46	33	20	0	44	32	24	84	16	0	0	72	28	0	0
		1000	0	49	29	22	0	32	36	32	22	78	0	0	8	92	0	0
		2000	0	42	36	22	0	36	44	20	0	100	0	0	0	100	0	0
	1.5	200	11	45	29	15	8	32	40	20	100	0	0	0	100	0	0	0
		500	1	42	34	24	0	40	20	40	86	14	0	0	76	24	0	0
		1000	0	36	37	28	0	44	44	12	24	76	0	0	8	92	0	0
		2000	0	21	40	39	0	32	48	20	0	100	0	0	0	100	0	0
LQ	0.4	200	77	16	5	2	46	27	16	11	100	0	0	0	99	1	0	0
		500	71	19	7	3	34	31	16	19	100	0	0	0	100	0	0	0
		1000	63	24	8	6	31	28	19	22	100	0	0	0	100	0	0	0
		2000	47	32	16	6	17	32	30	21	100	0	0	0	100	0	0	0
	0.9	200	75	18	5	3	47	24	14	15	100	0	0	0	100	0	0	0
		500	67	22	7	4	34	30	16	20	100	0	0	0	100	0	0	0
		1000	56	29	10	6	36	20	23	21	100	0	0	0	100	0	0	0
		2000	40	33	18	10	16	34	24	26	100	0	0	0	100	0	0	0
	1.5	200	71	18	9	3	50	24	12	14	100	0	0	0	100	0	0	0
		500	60	24	11	5	48	28	12	12	100	0	0	0	100	0	0	0
		1000	43	36	15	7	30	26	22	22	100	0	0	0	100	0	0	0
		2000	32	36	22	11	22	24	32	22	100	0	0	0	100	0	0	0

Table B.12: Proportion of times a given model is selected based on the lowest AIC and CAIC value for Population B Split 2.The shaded columns represent the true number of latent classes in the population.

			AIC								CAIC							
			No Covariates				Covariates				No Covariates				Covariates			
Quality	Effect	Sample	2	3	4	5	2	3	4	5	2	3	4	5	2	3	4	5
HQ	0.4	200	0	54	33	13	0	45	37	18	0	100	0	0	0	100	0	0
		500	0	48	36	16	0	40	34	26	0	100	0	0	0	100	0	0
		1000	0	43	38	19	0	34	38	28	0	100	0	0	0	100	0	0
		2000	0	43	42	15	0	36	40	24	0	100	0	0	0	100	0	0
	0.9	200	0	51	38	12	0	43	37	20	0	100	0	0	0	100	0	0
		500	0	45	37	18	0	40	36	25	0	100	0	0	0	100	0	0
		1000	0	43	36	21	0	33	40	28	0	100	0	0	0	100	0	0
		2000	0	40	40	20	0	37	37	26	0	100	0	0	0	100	0	0
	1.5	200	0	51	36	14	0	43	35	22	0	100	0	0	0	100	0	0
		500	0	41	38	21	0	38	38	24	0	100	0	0	0	100	0	0
		1000	0	37	39	24	0	34	39	27	0	100	0	0	0	100	0	0
		2000	0	26	40	33	0	37	36	27	0	100	0	0	0	100	0	0
MQ	0.4	200	0	44	34	21	0	30	36	33	43	57	0	0	34	66	0	0
		500	0	47	31	22	0	36	33	31	0	100	0	0	0	100	0	0
		1000	0	45	34	21	0	36	32	31	0	100	0	0	0	100	0	0
		2000	0	43	33	24	0	31	36	32	0	100	0	0	0	100	0	0
	0.9	200	0	44	36	19	0	31	36	33	45	55	0	0	33	67	0	0
		500	0	43	35	22	0	33	36	31	0	100	0	0	0	100	0	0
		1000	0	45	32	23	0	36	33	31	0	100	0	0	0	100	0	0
		2000	0	38	39	23	0	33	35	32	0	100	0	0	0	100	0	0
	1.5	200	0	43	33	24	0	33	35	32	48	52	0	0	36	64	0	0
		500	0	38	35	27	0	33	33	34	0	100	0	0	0	100	0	0
		1000	0	33	36	32	0	31	36	32	0	100	0	0	0	100	0	0
		2000	0	17	40	43	0	34	32	34	0	100	0	0	0	100	0	0
LQ	0.4	200	72	19	7	3	40	26	21	13	100	0	0	0	100	0	0	0
		500	59	23	12	5	30	23	21	26	100	0	0	0	100	0	0	0
		1000	45	35	14	7	23	27	23	26	100	0	0	0	100	0	0	0
		2000	32	37	18	13	17	27	26	29	100	0	0	0	100	0	0	0
	0.9	200	69	21	7	3	37	27	22	14	100	0	0	0	100	0	0	0
		500	56	26	11	7	26	26	23	25	100	0	0	0	100	0	0	0
		1000	42	33	18	7	21	27	25	27	100	0	0	0	100	0	0	0
		2000	33	33	22	11	17	31	22	30	100	0	0	0	100	0	0	0
	1.5	200	65	21	10	5	37	27	22	14	100	0	0	0	100	0	0	0
		500	51	29	13	8	26	26	23	25	100	0	0	0	100	0	0	0
		1000	38	33	18	11	21	27	25	27	100	0	0	0	100	0	0	0
		2000	25	33	26	16	17	31	22	30	100	0	0	0	100	0	0	0

Table B.13: Proportion of times a given model is selected based on the lowest BIC and aBIC value for Population A Split 1. The shaded columns represent the true number of latent classes in the population.

			BIC								aBIC							
			No Covariates				Covariates				No Covariates				Covariates			
Quality	Effect	Sample	2	3	4	5	2	3	4	5	2	3	4	5	2	3	4	5
HQ	0.4	200	58	42	0	0	61	39	0	0	1	73	24	3	2	69	24	6
		500	4	97	0	0	5	95	0	0	0	100	0	0	0	99	2	0
		1000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
		2000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
	0.9	200	32	69	0	0	29	72	0	0	1	71	24	5	1	65	28	7
		500	0	100	0	0	0	100	0	0	0	99	1	0	0	98	3	0
		1000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
		2000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
	1.5	200	5	95	0	0	2	98	0	0	0	68	28	4	0	66	28	7
		500	0	100	0	0	0	100	0	0	0	99	2	0	0	99	2	0
		1000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
		2000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
MQ	0.4	200	100	1	0	0	100	1	0	0	21	54	19	7	19	53	21	8
		500	89	12	0	0	91	10	0	0	19	82	0	0	20	80	1	0
		1000	37	64	0	0	44	57	0	0	2	98	0	0	2	98	0	0
		2000	2	99	0	0	2	99	0	0	0	100	0	0	0	100	0	0
	0.9	200	98	3	0	0	96	4	0	0	13	55	25	8	9	52	29	12
		500	71	30	0	0	62	38	0	0	6	94	1	0	3	95	3	0
		1000	6	95	0	0	4	96	0	0	0	100	0	0	0	100	0	0
		2000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
	1.5	200	87	14	0	0	60	40	0	0	4	62	27	7	1	53	31	16
		500	22	79	0	0	3	97	0	0	0	98	3	0	0	96	4	0
		1000	0	100	0	0	0	100	0	0	0	99	1	0	0	100	1	0
		2000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
LQ	0.4	200	100	0	0	0	100	0	0	0	88	10	2	1	69	18	7	7
		500	100	0	0	0	100	0	0	0	100	0	0	0	100	0	1	0
		1000	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0	0
		2000	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0	0
	0.9	200	100	0	0	0	100	0	0	0	86	11	2	1	68	18	8	7
		500	100	0	0	0	100	0	0	0	100	0	0	0	100	0	1	0
		1000	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0	0
		2000	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0	0
	1.5	200	100	0	0	0	100	0	0	0	85	11	3	2	61	20	12	8
		500	100	0	0	0	100	0	0	0	100	0	0	0	98	3	0	0
		1000	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0	0
		2000	100	0	0	0	100	0	0	0	100	0	0	0	100	1	0	0

Table B.14: Proportion of times a given model is selected based on the lowest BIC and aBIC value for Population A Split 2. The shaded columns represent the true number of latent classes in the population.

			BIC								aBIC							
			No Covariates				Covariates				No Covariates				Covariates			
Quality	Effect	Sample	2	3	4	5	2	3	4	5	2	3	4	5	2	3	4	5
HQ	0.4	200	0	100	0	0	0	100	0	0	0	66	28	6	0	61	31	8
		500	0	100	0	0	0	100	0	0	0	100	0	0	0	99	1	0
		1000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
		2000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
	0.9	200	0	100	0	0	0	100	0	0	0	63	33	4	0	67	27	6
		500	0	100	0	0	0	100	0	0	0	99	1	0	0	100	0	0
		1000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
		2000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
	1.5	200	0	100	0	0	0	100	0	0	0	61	34	5	0	55	31	14
		500	0	100	0	0	0	100	0	0	0	99	1	0	0	99	1	0
		1000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
		2000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
MQ	0.4	200	35	65	0	0	42	58	0	0	0	59	29	12	0	50	37	13
		500	0	100	0	0	0	100	0	0	0	97	3	0	0	97	3	0
		1000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
		2000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
	0.9	200	23	77	0	0	15	85	0	0	0	61	33	6	0	52	35	13
		500	0	100	0	0	0	100	0	0	0	99	1	0	0	99	1	0
		1000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
		2000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
	1.5	200	11	89	0	0	2	98	0	0	0	62	30	8	0	53	31	16
		500	0	100	0	0	0	100	0	0	0	98	2	0	0	97	3	0
		1000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
		2000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
LQ	0.4	200	100	0	0	0	100	0	0	0	81	12	4	3	51	29	14	6
		500	100	0	0	0	100	0	0	0	100	0	0	0	98	2	0	0
		1000	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0	0
		2000	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0	0
	0.9	200	100	0	0	0	100	0	0	0	85	8	6	1	54	26	12	8
		500	100	0	0	0	100	0	0	0	100	0	0	0	97	3	0	0
		1000	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0	0
		2000	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0	0
	1.5	200	100	0	0	0	100	0	0	0	78	14	5	3	46	25	18	11
		500	100	0	0	0	100	0	0	0	99	1	0	0	97	3	0	0
		1000	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0	0
		2000	100	0	0	0	100	0	0	0	100	0	0	0	99	1	0	0

Table B.15: Proportion of times a given model is selected based on the lowest BIC and aBIC value for Population B Split 1.The shaded columns represent the true number of latent classes in the population.

			BIC								aBIC							
			No Covariates				Covariates				No Covariates				Covariates			
Quality	Effect	Sample	2	3	4	5	2	3	4	5	2	3	4	5	2	3	4	5
HQ	0.4	200	30	70	0	0	27	73	0	0	1	73	21	6	0	75	22	3
		500	1	99	0	0	1	99	0	0	0	99	1	0	0	100	0	0
		1000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
		2000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
	0.9	200	30	71	0	0	29	71	0	0	1	72	25	3	0	70	28	2
		500	1	99	0	0	1	99	0	0	0	99	1	0	0	99	1	0
		1000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
		2000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
	1.5	200	32	69	0	0	28	72	0	0	1	68	27	4	0	66	29	5
		500	1	99	0	0	1	99	0	0	0	98	3	0	0	99	1	0
		1000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
		2000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
MQ	0.4	200	97	3	0	0	96	4	0	0	16	57	20	7	16	44	32	8
		500	68	32	0	0	52	48	0	0	6	94	1	0	0	100	0	0
		1000	11	90	0	0	0	100	0	0	0	100	0	0	0	100	0	0
		2000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
	0.9	200	97	3	0	0	96	4	0	0	15	54	23	8	12	56	20	12
		500	69	32	0	0	52	48	0	0	5	95	0	0	0	100	0	0
		1000	10	90	0	0	0	100	0	0	0	100	0	0	0	100	0	0
		2000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
	1.5	200	98	3	0	0	100	0	0	0	16	54	22	9	8	48	32	12
		500	71	29	0	0	52	48	0	0	6	92	2	0	0	100	0	0
		1000	12	88	0	0	0	100	0	0	0	100	0	0	0	100	0	0
		2000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
LQ	0.4	200	100	0	0	0	99	1	0	0	86	11	2	1	64	22	13	1
		500	100	0	0	0	100	0	0	0	100	0	0	0	99	1	0	0
		1000	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0	0
		2000	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0	0
	0.9	200	100	0	0	0	100	0	0	0	86	11	3	1	62	22	12	4
		500	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0	0
		1000	100	0	0	0	100	0	0	0	100	0	0	0	99	1	0	0
		2000	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0	0
	1.5	200	100	0	0	0	100	0	0	0	82	13	4	2	60	28	4	8
		500	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0	0
		1000	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0	0
		2000	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0	0

Table B.16: Proportion of times a given model is selected based on the lowest BIC and aBIC value for Population B Split 2. The shaded columns represent the true number of latent classes in the population.

			BIC								aBIC							
			No Covariates				Covariates				No Covariates				Covariates			
Quality	Effect	Sample	2	3	4	5	2	3	4	5	2	3	4	5	2	3	4	5
HQ	0.4	200	0	100	0	0	0	100	0	0	0	68	26	7	0	61	30	10
		500	0	100	0	0	0	100	0	0	0	98	2	0	0	98	2	0
		1000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
		2000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
	0.9	200	0	100	0	0	0	100	0	0	0	65	29	5	0	59	30	11
		500	0	100	0	0	0	100	0	0	0	98	2	0	0	99	1	0
		1000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
		2000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
	1.5	200	0	100	0	0	0	100	0	0	0	64	29	7	0	56	34	10
		500	0	100	0	0	0	100	0	0	0	98	2	0	0	98	2	0
		1000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
		2000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
MQ	0.4	200	22	78	0	0	14	86	0	0	0	58	31	11	0	50	33	17
		500	0	100	0	0	0	100	0	0	0	98	2	0	0	96	4	0
		1000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
		2000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
	0.9	200	23	77	0	0	14	86	0	0	0	58	31	11	0	48	34	18
		500	0	100	0	0	0	100	0	0	0	97	2	0	0	96	3	0
		1000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
		2000	0	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0
	1.5	200	24	76	0	0	14	86	0	0	0	57	32	11	0	49	33	18
		500	0	100	0	0	0	100	0	0	0	96	3	0	0	98	2	0
		1000	0	100	0	0	0	100	0	0	0	99	1	0	0	100	0	0
		2000	0	100	0	0	0	100	0	0	0	99	1	0	0	100	0	0
LQ	0.4	200	100	0	0	0	100	0	0	0	83	14	2	1	58	25	12	5
		500	100	0	0	0	100	0	0	0	100	0	0	0	99	1	0	0
		1000	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0	0
		2000	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0	0
	0.9	200	100	0	0	0	100	0	0	0	81	14	4	1	57	23	15	5
		500	100	0	0	0	100	0	0	0	99	1	0	0	97	3	0	0
		1000	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0	0
		2000	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0	0
	1.5	200	100	0	0	0	100	0	0	0	76	18	5	1	57	23	15	5
		500	100	0	0	0	100	0	0	0	99	1	0	0	97	3	0	0
		1000	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0	0
		2000	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0	0

Bibliography

Agresti, A. (2002). *Categorical Data Analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. In E. Parzen, K. Tanabe, and G. Kitagawa, editors, *Selected Papers of Hirotugu Akaike*, pages 199–213. Springer New York, New York, NY.

Asparouhov, T. and Muthén, B. (2012). Using Mplus TECH11 and TECH14 to test the number of latent classes.

Baudry, J.-P., Raftery, A. E., Celeux, G., Lo, K., and Gottardo, R. (2010). Combining Mixture Components for Clustering. *Journal of Computational and Graphical Statistics*, **19**(2), 332–353.

Biernacki, C. and Govaert, G. (1997). Using the classification likelihood to choose the number of clusters. *Computing Science and Statistics*, **20**, 451–457.

Biernacki, C., Celeux, G., and Govaert, G. (2000). Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**(7), 719–725.

Bozdogan, H. (1987). Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, **52**(3), 345–370.

Chen, Q., Luo, W., Palardy, G. J., Glaman, R., and McEnturff, A. (2017). The Efficacy of Common Fit Indices for Enumerating Classes in Growth Mixture Models When Nested Data Structure Is Ignored: A Monte Carlo Study. *SAGE Open*, **7**(1), 215824401770045.

Chung, T., Pedersen, S. L., Kim, K. H., Hipwell, A. E., and Stepp, S. D. (2014). Racial Differences in Type of Alcoholic Beverage Consumed During Adolescence in the Pittsburgh Girls Study. *Alcoholism: Clinical and Experimental Research*, **38**(1), 285–293.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. L. Erlbaum Associates, Hillsdale, N.J, 2nd ed edition.

Collins, L. M. and Lanza, S. T. (2010). *Latent class and latent transition analysis with applications in the social, behavioral, and health sciences*. Wiley, Hoboken, NJ. OCLC: 1141423188.

Crow, S. J., Swanson, S. A., Peterson, C. B., Crosby, R. D., Wonderlich, S. A., and Mitchell, J. E. (2012). Latent class analysis of eating disorders: Relationship to mortality. *Journal of Abnormal Psychology*, **121**(1), 225–231.

Dayton, C. M. and Macready, G. B. (2012). Concomitant-Variable Latent-Class Models. *Journal of the American Statistical Association*.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum Likelihood from

Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, **39**(1), 1–22.

Diallo, T. M. O., Morin, A. J. S., and Lu, H. (2017a). The impact of total and partial inclusion or exclusion of active and inactive time invariant covariates in growth mixture models. *Psychological Methods*, **22**(1), 166–190.

Diallo, T. M. O., Morin, A. J. S., and Lu, H. (2017b). Performance of growth mixture models in the presence of time-varying covariates. *Behavior Research Methods*, **49**(5), 1951–1965.

Dias, J. G. (2006). Latent class analysis and model selection. In M. Spiliopoulou, R. Kruse, C. Borgelt, A. Nürnberger, and W. Gaul, editors, *From Data and Information Analysis to Knowledge Engineering*, pages 95–102, Berlin, Heidelberg. Springer Berlin Heidelberg.

Dziak, J. J., Lanza, S. T., and Tan, X. (2014). Effect Size, Statistical Power, and Sample Size Requirements for the Bootstrap Likelihood Ratio Test in Latent Class Analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, **21**(4), 534–552.

Dziak, J. J., Coffman, D. L., Lanza, S. T., Li, R., and Jermiin, L. S. (2018). Sensitivity and Specificity of Information Criteria. preprint, Bioinformatics.

Everitt, B. (1988). A Monte Carlo Investigation of the Likelihood Ratio Test for Number of Classes in Latent Class Analysis. *Multivariate Behavioral Research*, **23**(4), 531–538.

Everitt, B. S. (1984). *An Introduction to Latent Variable Models*. Springer Netherlands, Dordrecht.

Fonseca, J. R. and Cardoso, M. G. (2007). Mixture-model cluster analysis using information theoretical criteria. *Intelligent Data Analysis*, **11**(2), 155–173.

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

Haberman, S. J. (1974). Log-Linear Models for Frequency Tables with Ordered Classifications. *Biometrics*, **30**(4), 589–600.

Hagenaars, J. (1993). *Loglinear models with Latent Variables*. SAGE Publications, Inc., 2455 Teller Road, Thousand Oaks California 91320 United States of America.

Hagenaars, J. A. (1998). Categorical Causal Modeling: Latent Class Analysis and Directed Log-Linear Models with Latent Variables. *Sociological Methods & Research*, **26**(4), 436–486.

Hallquist, M. N. and Wiley, J. F. (2018). MplusAutomation: An R package for facilitating large-scale latent variable analyses in Mplus. *Structural Equation Modeling*, pages 1–18.

Haughton, D. M. A. (1988). On the Choice of a Model to Fit Data from an Exponential Family. *The Annals of Statistics*, **16**(1), 342–355.

Henson, J. M., Reise, S. P., and Kim, K. H. (2007). Detecting Mixtures From Structural Model Differences Using Latent Variable Mixture Modeling: A Comparison of Relative Model Fit Statistics. *Structural Equation Modeling: A Multidisciplinary Journal*, **14**(2), 202–226.

Hu, J., Leite, W. L., and Gao, M. (2017). An evaluation of the use of covariates to assist in class enumeration in linear growth mixture modeling. *Behavior Research Methods*, **49**(3), 1179–1190.

Huang, G.-H. and Bandeen-Roche, K. (2004). Building an identifiable latent class model with covariate effects on underlying and measured variables. *Psychometrika*, **69**(1), 5–32.

Jang, Y., Park, N. S., Chiriboga, D. A., and Kim, M. T. (2017). Latent Profiles of Acculturation and Their Implications for Health: A Study With Asian Americans in Central Texas. *Asian American journal of psychology*, **8**(3), 200–208.

Janssen, J. H. M., van Laar, S., de Rooij, M. J., Kuha, J., and Bakk, Z. (2019). The Detection and Modeling of Direct Effects in Latent Class Analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, **26**(2), 280–290.

Jeffries, N. O. (2003). A note on 'Testing the number of components in a normal mixture'. *Biometrika*, **90**(4), 991–994.

Lazarsfeld, P. F. and Henry, N. W. (1968). *Latent structure analysis*. Houghton, Mifflin, New York. OCLC: 433623.

Li, L. and Hser, Y.-I. (2011). On Inclusion of Covariates for Class Enumeration of Growth Mixture Models. *Multivariate Behavioral Research*, **46**(2), 266–302.

Lo, Y. (2001). Testing the number of components in a normal mixture. *Biometrika*, **88**(3), 767–778.

Lubke, G. and Muthén, B. O. (2007). Performance of Factor Mixture Models as a

Function of Model Size, Covariate Effects, and Class-Specific Parameters. *Structural Equation Modeling: A Multidisciplinary Journal*, **14**(1), 26–47.

Lukočienė, O., Varriale, R., and Vermunt, J. K. (2010). 6. The Simultaneous Decision(s) about the Number of Lower- and Higher-Level Classes in Multilevel Latent Class Analysis. *Sociological Methodology*, **40**(1), 247–283.

Masyn, K. E. (2013). *Latent Class Analysis and Finite Mixture Modeling*. Oxford University Press.

Masyn, K. E. (2017). Measurement Invariance and Differential Item Functioning in Latent Class Analysis With Stepwise Multiple Indicator Multiple Cause Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, **24**(2), 180–197.

Masyn, K. E., Petras, H., and Liu, W. (2014). Growth Curve Models with Categorical Outcomes. In G. Bruinsma and D. Weisburd, editors, *Encyclopedia of Criminology and Criminal Justice*, pages 2013–2025. Springer New York, New York, NY.

McCullagh, P. (2005). Proportional-Odds Model. In P. Armitage and T. Colton, editors, *Encyclopedia of Biostatistics*, page b2a10049. John Wiley & Sons, Ltd, Chichester, UK.

McLachlan, G. and Ng, S. (2000). A comparison of some information criteria for the number of components in a mixture model. Technical Report, Department of Mathematics, University of Queensland, Brisbane.

McLachlan, G. J. (1987). On Bootstrapping the Likelihood Ratio Test Stastistic for the Number of Components in a Normal Mixture. *Applied Statistics*, **36**(3), 318.

McLachlan, G. J. and Peel, D. (2000). *Finite mixture models*. Wiley series in probability and statistics. Applied probability and statistics section. Wiley, New York.

Morgan, G. B. (2015). Mixed Mode Latent Class Analysis: An Examination of Fit Index Performance for Classification. *Structural Equation Modeling: A Multidisciplinary Journal*, **22**(1), 76–86.

Morgan, G. B., Hodge, K. J., and Baggett, A. R. (2016). Latent profile analysis with nonnormal mixtures: A Monte Carlo examination of model selection using fit indices. *Computational Statistics & Data Analysis*, **93**, 146–161.

Morovati, D. (2014). *The Intersection of Sample Size, Number of Indicators, and Class Enumeration in LCA: A Monte Carlo Study*. Ph.D. thesis, UC Santa Barbara.

Muthén, B. O. (2008). Latent variable hybrids: Overview of old and new models. In G. R. Hancock and K. M. Samuelsen, editors, *Advances in latent variable mixture models*, pages 1–24. Information Age Publishing, Inc., Charlotte, NC. OCLC: ocn171287678.

Muthén, L. and Muthén, B. .-. (1997-2017). *Mplus User's Guide. Eighth Edition.* Muthén & Muthén, Los Angeles, CA.

Muthén, L. K. and Muthén, B. (2002). How To Use A Monte Carlo Study To Decide On Sample Size and Determine Power.

Nylund, K. L., Asparouhov, T., and Muthén, B. O. (2007). Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. *Structural Equation Modeling: A Multidisciplinary Journal*, **14**(4), 535–569.

Nylund-Gibson, K. and Choi, A. Y. (2018). Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science*, **4**(4), 440–461.

Nylund-Gibson, K. and Masyn, K. E. (2016). Covariates and Mixture Modeling: Results of a Simulation Study Exploring the Impact of Misspecified Effects on Class Enumeration. *Structural Equation Modeling: A Multidisciplinary Journal*, **23**(6), 782–797.

Park, J. and Yu, H.-T. (2018). A Comparison of Approaches for Estimating Covariate Effects in Nonparametric Multilevel Latent Class Models. *Structural Equation Modeling: A Multidisciplinary Journal*, **25**(5), 778–790.

Peugh, J. and Fan, X. (2015). Enumeration Index Performance in Generalized Growth Mixture Models: A Monte Carlo Test of Muthén's (2003) Hypothesis. *Structural Equation Modeling: A Multidisciplinary Journal*, **22**(1), 115–131.

Qu, Y., Tan, M., and Kutner, M. H. (1996). Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics*, **52**(3), 797–810.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramaswamy, V., Desarbo, W. S., Reibstein, D. J., and Robinson, W. T. (1993). An Empirical Pooling Approach for Estimating Marketing Mix Elasticities with PIMS Data. *Marketing Science*, **12**(1), 103–124.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, **6**(2).

Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, **52**(3), 333–343.

Tekle, F. B., Gudicha, D. W., and Vermunt, J. K. (2016). Power analysis for the bootstrap likelihood ratio test for the number of classes in latent class models. *Advances in Data Analysis and Classification*, **10**(2), 209–224.

Tofighi, D., Enders, C., and Samuelsen, K. (2008). Identifying the correct number of classes in growth mixture models. In G. Hancock, editor, *Advances in Latent Variable Mixture Models*, pages 317–341. Information Age Publishing, Inc, Charlotte, NC.

Vermunt, J. and Magidson, J. (2016). *Upgrade Manual for Latent GOLD 5.1*. Statistical Innovations Inc., Belmont,Massachusetts.

Vermunt, J. K. (2003). Multilevel Latent Class Models. *Sociological Methodology*, **33**(1), 213–239.

Vermunt, J. K. (2008). Latent class and finite mixture models for multilevel data sets. *Statistical Methods in Medical Research*, **17**(1), 33–51.

Vuong, Q. H. (1989). Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica*, **57**(2), 307.

White, A. and Murphy, T. B. (2014). **BayesLCA** : An *R* Package for Bayesian Latent Class Analysis. *Journal of Statistical Software*, **61**(13).

Whittaker, T. A. and Miller, J. E. (2021). Exploring the Enumeration Accuracy of Cross-Validation Indices in Latent Class Analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, **28**(3), 376–390.

Woodroofe, M. (1982). On Model Selection and the ARC Sine Laws. *The Annals of Statistics*, **10**(4), 1182–1194.

Yang, C.-C. and Yang, C.-C. (2007). Separating Latent Classes by Information Criteria. *Journal of Classification*, **24**(2), 183–203.