Statistics and Its Interface Volume 0 (2019) 1–4

A latent class selection model for categorical response variables with nonignorably missing data*

Jung Wun Lee, Ofer Harel†

We develop a new selection model for nonignorable missing values in multivariate categorical response variables by assuming that the response variables and their missingness can be summarized into categorical latent variables. Our proposed model contains two categorical latent variables. One latent variable summarizes the response patterns while the other describes the response variables' missingness. Our selection model is an alternative method to other incomplete data methods when the incomplete data mechanism is nonignorable. We implement simulation studies to evaluate the performance of the proposed method and analyze the General Social Survey 2018 data to demonstrate its performance.

AMS 2000 subject classifications: Primary 00K00, 00K01; secondary 00K02.
Keywords and phrases: Latent class model, Selection model, Missing not at random, EM algorithm, Bayesian inference.

1. INTRODUCTION

Nonresponse is a common but serious issue in data collection and analysis especially in survey-based research. Observations with any nonresponse increase the risk of obtaining inaccurate inferences on parameters of interest because it may degrade the performance of confidence intervals, reduce the statistical power, and magnify biases in parameter estimates. In this sense, appropriate adjustments to nonresponse are needed for research with incomplete data such as surveys or other observational studies.

Handling nonresponse generally requires different techniques depending on the types of nonresponse and the assumptions the researchers are willing to make on the missingness process. The three common nonresponse mechanisms are MAR (missing at random), MCAR (missing completely at random), and MNAR (missing not at random). MAR denotes that the missingness on a variable does not depend on itself but may depend on the other observed variables in the data. MCAR is a special case of MAR where missingness on a variable is not related to any observed or missing variables. MNAR occurs if the missingness depends on unobserved values of the data.

Since different nonresponse mechanisms require different remedies to reduce biases and allow for efficient estimates, it is important to apply appropriate technique in dealing with nonresponse. For example, Demirtas and Schafer [1] showed that applying MAR-based methods or using inappropriate MNAR-based methods on a MNAR data may cause biased parameter estimates which reduces the coverage of their confidence intervals. To overcome such potential problems, we are motivated to develop new methodology under a MNAR condition that can handle nonignorable nonresponse and is less sensitive to the choice of the nonignorable model compared to other selection models.

Developing appropriate tools for missing values under MNAR is important for prospective analysts because these are common problems in social and clinical sciences such as education [2], psychology [3], and clinical trial research [4]. For example, self-reported questionnaires are commonly used in research on substance use, risk behavior, social attitudes, or smoking. In these surveys for example, some respondents might refuse to report their true behaviors because they don't want to reveal their behaviors or opinions which may be considered undesirable. Also, some respondents who do not disclose may dismiss some items because they do not think these items apply to them.

Three general types of methods have been suggested for incomplete data under MNAR. Selection models describe the conditional distribution of the missingness given the complete data [5, 6]. Another possible approach to the MNAR mechanism is a pattern mixture model which decomposes the joint distribution of complete data and missingness into the conditional distribution of complete data given the missingness, and the marginal distribution of missingness [7]. The third method is referred to as a shared parameter model which assumes the existence of a latent variable that explains the association between the nonresponse patterns and the observed variables [8]. In this article, we are interested in developing a new selection model because we are interested in individuals' propensities to non-respond. A selection model may answer this question as it models how an individual's response pattern may impact its probability

arXiv: 0000.0000
*This material is partially based upon work supported by the National Science Foundation under Grant No. DMS-2218847. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.
†Corresponding author.

of being observed or missing. In this sense, our goal is to use a latent class framework in modeling missing patterns and complete data and to avoid the instability and extreme sensitivity of conventional selection models, as pointed in [9, 10, 11]. Because our proposed model explains the impact of complete data on missingness via latent variables, it can also be considered as a combination of selection model and shared-parameter model.

Several papers suggested summarizing a large number of missingness patterns into few representative prototypes via latent class model [12]. For example, latent classes are used for summarizing mixture patterns in a pattern mixture model [13] or reducing the number of missingness patterns in dropouts [14]. Each of these models adopted the shared parameter model whose response variables followed the normal distribution, and the latent classes were defined by the missingness indicators. Jung et al. [10] suggested a latent class selection model for nonignorable missing values by constructing a categorical latent variable using missing indicators of response variables, while treating the response variables and other variables as covariates.

The goal of this paper is to suggest a latent class selection model which can deal with non-ignorably missing values by imposing a latent igorability assumption as suggested in [15, 16]. Latent ignorability assumes that the missingness depends on a summary of the missing values rather than all the missing values themselves. In this paper, we posit a categorical latent variable which summarizes the patterns of missingness. In Section 2, we briefly review a latent class model (LCM) and missingness mechanisms to propose our new latent class selection model (LCSM) for non-ignorable missing data. In Section 3, we describe two estimation strategies for LCSM. In Section 4 we apply the LCSM to the General Social Study (GSS) data to illustrate that missing patterns can be modeled by a latent class structure. Conclusions and directions for further research are discussed in Section 5.

2. MODEL

Latent class model

A latent class model postulates a categorical latent variable which provides a set of partitions of population that cannot be measured directly via response variables. Let $\mathbf{Y}_i = [Y_{i1}, \ldots, Y_{iP}]$ be discrete random variables that have $r_1, \ldots, r_P$ categories, respectively, and $\mathbf{X}_i = [X_{i1}, \ldots, X_{iK}]$ are individual covariates such as demographic factors. The number of possible response patterns of $\mathbf{Y}_i$ is $\prod_{j=1}^{P} r_j$, but it can be summarized into S number of patterns by introducing a discrete latent variable U that has S categories. Individuals with the same latent class membership $U = u$ are homogeneous in their responses $\mathbf{Y}$, and individuals in different classes will show different response patterns. In this manner, we reduce a $\prod_{j=1}^{P} r_j$ dimensional contingency table into S categories. Further, we assume that the $[Y_{i1}, \ldots, Y_{iP}]$ are conditionally independent when U is observed. Namely, this is a local independence assumption [17] in that all existing associations among $[Y_{i1}, \ldots, Y_{iP}]$ can be explained by a latent variable U. Under such assumptions, we can construct the joint distribution of $\mathbf{Y}_i$ as follows:

$$P(\mathbf{Y}_i \mid \mathbf{X}_i) = \sum_{u=1}^{S} P(U = u \mid \mathbf{X}_i) \prod_{p=1}^{P} P(Y_{ip} \mid U = u). \quad (1)$$

Here, the non-differential measurement assumption [17] has been imposed such that $P(Y_{ip} \mid U, \mathbf{X}_i) = P(Y_{ip} \mid U)$ for all $p = 1 \ldots P$. This means that the conditional distribution of $\mathbf{Y}_i$ given U are invariant to the covariates $\mathbf{X}_i$. If such assumption does not hold, then $P(Y_{ip} \mid U, \mathbf{X}_i)$ needs to be modeled via additional parameterizations such as multinomial logistic regression [18].

Missingness mechanisms

Suppose the data $\mathbf{Y} \in \mathbb{R}^{N \times P}$ consists of N observations with P random variables with some missing values. We denote $\mathbf{Y}_i = [Y_{i1}, \ldots, Y_{iP}]$ to be the response vector of the ith individual. Next, we define a matrix of random variables $\mathbf{R} \in \mathbb{R}^{N \times P}$, $i = 1 \ldots, N$, $p = 1 \ldots, P$, where $R_{ip} = 1$ if Y_{ip} was observed and $R_{ip} = 0$ if Y_{ip} was missing. Note that R_{ip} is a random variable because the missingness on Y_{ip} is random. Next, we can decompose the ith response vector into $[\mathbf{Y}_i^O, \mathbf{Y}_i^m]$, where $\mathbf{Y}_i^O$ and $\mathbf{Y}_i^m$ are notations for observed and missing individual outcomes, respectively. Let θ be parameters related to the data $\mathbf{Y}$ and ξ be parameters related to missingness indicator $\mathbf{R}$. The complete data likelihood of ith individual is equivalent to the joint distribution of $[\mathbf{Y}_i^O, \mathbf{Y}_i^m, \mathbf{R}_i \mid \theta, \xi]$. Note that $\mathbf{R}_i$ are completely observed because it should be clear whether each record Y_{ip} is observed or missing. Now, we can decompose the complete data likelihood as follows.

$$P(\mathbf{Y}_i^O, \mathbf{Y}_i^m, \mathbf{R}_i \mid \theta, \xi) = P(\mathbf{R}_i \mid \mathbf{Y}_i^O, \mathbf{Y}_i^m, \theta, \xi) \times P(\mathbf{Y}_i^O, \mathbf{Y}_i^m \mid \theta, \xi). \quad (2)$$

The complete data likelihood can be written as the product of the conditional probability of $\mathbf{R}_i$ given $\mathbf{Y}_i$ and the marginal distribution of $\mathbf{Y}_i$ (see Eq. 2). The conditional probability of $\mathbf{R}_i$ given $\mathbf{Y}_i$ can be referred to as the missingness mechanism which describes the probabilistic occurrence of missing values [7, 15, 6]. The missingness mechanism is MAR if the conditional distribution of $\mathbf{R}$ given $\mathbf{Y}$ does not depend on $\mathbf{Y}^m$, that is, $P(\mathbf{R} \mid \mathbf{Y}) = P(\mathbf{R} \mid \mathbf{Y}^O)$. MCAR is a special case of MAR in that the missingness pattern R is independent of $\mathbf{Y}$ and thus $P(\mathbf{R} \mid \mathbf{Y}) = P(\mathbf{R})$. Further, we say the missingness mechanism is ignorable if the missingness mechanism is MAR (or MCAR) (i) and (ii) the joint parameter space of $[\xi, \theta]$ is the Cartesian product of the respective

parameter space of ξ and θ. That is, $\Omega_{\xi\times\theta} = \Omega_\xi \times \Omega_\theta$. When at least one of these two conditions does not hold, then the missingness mechanism becomes non-ignorable (MNAR).

In practice, we are interested in inferences of the data parameters θ based on the observed data likelihood. The observed likelihood of each individual will consist of $[\mathbf{Y}_i^O, \mathbf{R}_i]$ and can be obtained by integrating the complete data likelihood with respect to $\mathbf{Y}_i^m$. Now, when the missingness mechanism is ignorable, the contribution of $\mathbf{R}$ on the observed likelihood can be excluded from the estimation for data parameter θ [6]. On the other hand, the conditional distribution of $\mathbf{R}$ given $\mathbf{Y}$ must be included in the estimation if the ignorablity assumption does not hold.

Latent class selection model (LCSM)

Suppose the ignorability assumption does not hold and hence we should take $\mathbf{R}$ into account when estimating the data parameter θ. In this paper, we use a latent class structure to specify the nonresponse patterns $P(\mathbf{R} \mid \mathbf{Y})$ by summarizing the missingness patterns into several common categories. Let W denote a latent response propensity which determines the values of $[R_1, \ldots, R_P]$. Namely, W is a categorical latent variable identified through the missingness indicators $[R_1, \ldots, R_P]$ which identifies the groups of individuals with respect to their nonresponse patterns. We assume that the response patterns of $\mathbf{R}$ (i.e., the nonresponse patterns) are directly dependent on $[U, W]$, where U is another categorical latent variable measured by the response variables $\mathbf{Y} = [\mathbf{Y}^O, \mathbf{Y}^m]$. Fig.1 identifies the structure of our proposed LCSM model.

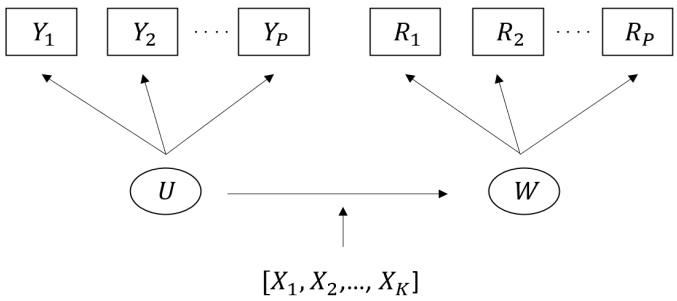

Figure 1. Structure of a latent class selection model

To construct the complete data likelihood of the model, we are assuming additional conditions as follows:

1. $[Y_1, \ldots, Y_P]$ and $[R_1, \ldots, R_P]$ are related only through U and W.
2. $[Y_1, \ldots, Y_P]$ and W are conditionally independent on U.
3. $[R_1, \ldots, R_P]$ and U are conditionally independent on W.

These assumptions imply that all the associations between the response variable $\mathbf{Y}$ and $\mathbf{R}$ can be explained by a latent variable U and a missingness propensity W. These assumptions are plausible in that they only assume the existence of two categorical latent variables and do not require additional assumptions, for example, a functional form of association between $[U, W]$. In this sense, these assumptions can be considered as the general case of other assumptions that require more specific details on the missing mechanism on $\mathbf{Y}$. In addition, the proposed model may yield better interpretations on the representative patterns of nonresponses as well as response patterns by having two seperate latent variable, rather than a single latent variable U. Based on these assumptions, the complete data likelihood of the model consists of $[\mathbf{R}, \mathbf{Y}^O, \mathbf{Y}^m, U, W]$, and the observed data likelihood can be obtained by integrating the complete data likelihood over $[\mathbf{Y}^m, U, W]$ as follows.

$$
\begin{aligned}
P(\mathbf{R}, \mathbf{Y}^O) &= \int\int\int P(\mathbf{R}, \mathbf{Y}^O, \mathbf{Y}^m, U, W) d\mathbf{Y}^m dU dW \\
&= \int\int\int P(\mathbf{R}, W \mid \mathbf{Y}^O, \mathbf{Y}^m, U) P(\mathbf{Y}^O, \mathbf{Y}^m, U) d\mathbf{Y}^m dU dW \\
&= \int\int\int P(\mathbf{R}, W \mid U) P(\mathbf{Y}^O \mid \mathbf{Y}^m, U) P(\mathbf{Y}^m, U) d\mathbf{Y}^m dU dW \\
&= \int\int\int P(\mathbf{R}, W \mid U) P(\mathbf{Y}^O \mid U) P(\mathbf{Y}^m, U) d\mathbf{Y}^m dU dW \\
&= \int\int\int P(\mathbf{R}, W \mid U) P(\mathbf{Y}^O \mid U) P(\mathbf{Y}^m \mid U) P(U) d\mathbf{Y}^m dU dW \\
&= \int\int P(\mathbf{R}, W \mid U) P(\mathbf{Y}^O \mid U) P(U) \int P(\mathbf{Y}^m \mid U) d\mathbf{Y}^m dU dW \\
&= \int\int P(\mathbf{R}, W \mid U) P(\mathbf{Y}^O \mid U) P(U) dU dW \qquad (3)
\end{aligned}
$$

The missingness mechanism $P(\mathbf{R}, W \mid U)$ is MNAR because $[\mathbf{R}, W]$ depend on U and U is the summary of $\mathbf{Y}$. In this way, the non-response indicator $\mathbf{R}$ is affected by both $\mathbf{Y}^O$ and $\mathbf{Y}^m$ via the latent variable U and missing propensity W. Harel and Schafer [15] described such mechanisms as *latent ignorability*, where missingness mechanism becomes ignorable if latent variables $[U, W]$ are observed. Further, we impose additional conditions on $\mathbf{Y}$ and $\mathbf{R}$ as follows

1. $[Y_1, \ldots, Y_P]$ are conditionally independent on U. That is, $P(Y_1, \ldots, Y_P \mid U) = \prod_{m=1}^{P} P(Y_m \mid U)$.
2. $[R_1, \ldots, R_P]$ are conditionally independent on U and W. That is, $P(R_1, \ldots, R_P \mid U, W) = \prod_{m=1}^{P} P(R_m \mid U, W)$.

These two assumptions allows the response variables $[Y_1, \ldots, Y_P]$ be conditionally independent given value of U, and missingness indicators $[R_1, \ldots, R_P]$ are conditionally independent given values of $[U, W]$. Now, we define four parameter types that constitute the proposed LCA model with non-ignorable missing values.

1. $\rho_{mk|u} = P(Y_m = k \mid U = u)$ is the measurement parameter for response variable Y_m given $U = u$, $k = 1, \ldots, r_m$, $m = 1 \ldots P$.

2. $\gamma_u = P(U = u)$ is the proportion of latent class $U = u$.
3. $\phi_{ph|w,u} = P(R_p = h \mid W = w, U = u)$ is the measurement parameter for non-response variable R_p given that $[U, W] = [u, w]$, $h = 0, 1$, $p = 1 \dots P$.
4. $\delta_{w|u} = P(W = w \mid U = u)$ is the proportion of latent missing propensity w, given that latent class membership is u.

Based on these four parameter types, we can formulate Eq. (4) in parameterized form as follows:

$$
\begin{aligned}
L_i &= \int\int P(\mathbf{Y}_i^O, \mathbf{Y}_i^m, \mathbf{R}_i, U, W \mid \mathbf{X}_i)\; d\mathbf{Y}^m dU dW \\
&= \int\int P(\mathbf{R}_i, W \mid U) P(\mathbf{Y}_i^O \mid U) P(U \mid \mathbf{X}_i)\; dU dW \\
&= \int\int P(\mathbf{R}_i \mid W, U) P(W \mid U) P(\mathbf{Y}_i^O \mid U) P(U \mid \mathbf{X}_i)\; dU dW \\
&= \int\int P(W \mid U) \prod_{p=1}^{P} P(R_{ip} \mid W, U) P(U \mid \mathbf{X}_i) \times \\
&\quad \times \prod_{m \in O_i} P(Y_{im} \mid U)\; dU dW \\
&= \sum_{w=1}^{D} \sum_{u=1}^{S} \left\{ \delta_{w|u} \prod_{p=1}^{P} \prod_{h=0}^{1} \phi_{ph|w,u}^{I(R_{ip}=h)} \gamma_u(\mathbf{X}_i) \prod_{m \in O_i} \prod_{k=1}^{r_m} \rho_{mk|u}^{I(Y_{im}=k)} \right\}.
\end{aligned} \tag{4}
$$

We assume that the covariates $\mathbf{X}_i$ only affects the distribution of U. Then, the effect of $\mathbf{X}_i$ on the prevalence of latent classes γ_u, $u = 1, \dots, S$ can be explained via baseline multinomial logistic regression. This assumption simplifies the model and provides a direct interpretation for the effect of $\mathbf{X}_i$ on U. Let $\boldsymbol{\beta}_u = [\beta_{1u}, \dots, \beta_{Ku}]'$ be a vector of coefficients which represent the effect of $\mathbf{X}_i = [X_{i1}, \dots, X_{iK}]'$ on uth latent class. Then, the prevalence of latent classes can be written as follows:

$$
\gamma_u(\mathbf{X}_i) = \frac{exp(\mathbf{X}_i \boldsymbol{\beta}_u)}{\sum_{s=1}^{S} exp(\mathbf{X}_i \boldsymbol{\beta}_s)}, \quad u = 1, \dots, S. \tag{5}
$$

3. PARAMETER ESTIMATION

In this section, we discuss two parameter estimation strategies for LCSM that can be adopted depending on the analysis goal, sample size, and the existence of subjective information on the parameters. One method is maximum likelihood (ML) estimation for researchers who seek asymptotic properties of estimates. The other is Bayesian estimation which uses additional information by placing priors on the parameters.

Maximum likelihood estimation via EM algorithm

The Expectation-Maximization (EM) algorithm [19] was devised to calculate maximum likelihood estimates and allows missing values. Since LCSM faces two types of missing values that are (i) nonresponse on the response variables and (ii) unobservable latent class memberships, the conventional EM method is a good strategy to obtain ML estimates.

The conventional EM algorithm consists of two steps: (i) in the E-step, we calculate the expectation of the log-complete data likelihood $E(logL^*)$, and (ii) in the M-step, we calculate the updated parameter estimates by maximizing the expectation. In the E-step, we need to calculate the conditional expectation of the latent variables $[U, W]$ given $[\mathbf{R}, \mathbf{Y}, \mathbf{X}]$. Since both latent variables are categorical, $[U, W \mid \mathbf{R}, \mathbf{Y}, \mathbf{X}]$ follows a multinomial distribution, whose probabilities for each level are denoted by $\theta_{i(u,w)}$, $u = 1, \dots, S$ and $w = 1, \dots, D$. Consequently, the joint probability $\theta_{i(u,w)} = P(U = u, W = w \mid \mathbf{R}_i, \mathbf{Y}_i, \mathbf{X}_i)$ of the latent variables $[U, W]$ given the ith observed responses and covariates are defined as follows.

$$
\begin{aligned}
\theta_{i(w,u)} &= \frac{\delta_{w|u} \prod_{p=1}^{P} \prod_{h=0}^{1} \phi_{ph|w,u}^{I(R_{ip}=h)} \gamma_u(\mathbf{X}_i) \prod_{m \in O_i} \prod_{k=1}^{r_m} \rho_{mk|u}^{I(Y_{im}=k)}}{\sum_{u=1}^{S} \sum_{w=1}^{D} \left\{ \delta_{w|u} \prod_{p=1}^{P} \prod_{h=0}^{1} \phi_{ph|w,u}^{I(R_{ip}=h)} \gamma_u(\mathbf{X}_i) \prod_{m \in O_i} \prod_{k=1}^{r_m} \rho_{mk|u}^{I(Y_{im}=k)} \right\}} \\
\theta_{i(u)} &= \sum_{w=1}^{D} \theta_{i(w,u)} \;\text{ and }\; \theta_{i(w)} = \sum_{u=1}^{S} \theta_{i(w,u)},\; u = 1, \dots, S,\; w = 1, \dots, D.
\end{aligned} \tag{6}
$$

In the E-step, we calculate the expectation of the complete log-likelihood using Eq. (6) as follows:

$$
\begin{aligned}
logL_i^* &= I(W_i = w, U_i = u) log\delta_{w|u} \\
&\quad + \sum_{m=1}^{P} \sum_{h=0}^{1} I(W_i = w, U_i = u) I(R_{ip} = h) log\phi_{ph|w,u} \\
&\quad + I(U_i = u) I(R_{ip} = h) log\gamma_u \\
&\quad + \sum_{m \in O_i} \sum_{k=1}^{r_m} I(U_i = u) I(Y_{im} = k) log\rho_{mk|u}
\end{aligned} \tag{7}
$$

$$
\begin{aligned}
E(logL_i^*) &= \sum_{p=1}^{P} \sum_{h=0}^{1} \theta_{i(w,u)} I(R_{ip} = h) log\phi_{ph|w,u} \\
&\quad + \theta_{i(w,u)} log\delta_{w|u} + \theta_{i(u)} log\gamma_u \\
&\quad + \sum_{m \in O_i} \sum_{k=1}^{r_m} \theta_{i(u)} I(Y_{im} = k) log\rho_{mk|u}.
\end{aligned} \tag{8}
$$

Since we assume the hierarchical latent class structure on $P(R \mid Y)$, the complete data likelihood and its expectation can be written in forms of multinomial distribution as in Eq. (7). Then in the M-step, we maximize Eq. (7) using the Lagrange Multipliers under the following constraints; (i) $\sum_{u=1}^{S} \sum_{d=1}^{D} \theta_{i(w,u)} = 1$ and (ii) $\sum_{u=1}^{S} \theta_{i(u)} = 1$. The maximizer of Eq. (7) with respect to each parameters can be calculated as follows:

$$(9)\quad \hat{\rho}_{mk|u} = \frac{\sum_{i=1}^{N} \theta_{i(u)} I(Y_{im} = k)}{\sum_{i=1}^{N} \theta_{i(u)}}, \quad \hat{\gamma}_u = \frac{1}{N} \sum_{i=1}^{N} \theta_{i(u)}$$

$$\hat{\phi}_{ph|w,u} = \frac{\sum_{i=1}^{N} \theta_{i(w,u)} I(R_{ip} = h)}{\sum_{i=1}^{N} \theta_{i(w,u)}}, \quad \hat{\delta}_{w|u} = \frac{\sum_{i=1}^{N} \theta_{i(w,u)}}{\sum_{i=1}^{N} \theta_{i(u)}}.$$

Through iteration by alternating E- and M-steps, we can numerically obtain ML estimates for $[\boldsymbol{\delta}, \boldsymbol{\rho}, \boldsymbol{\gamma}, \boldsymbol{\phi}]$. When an individual-specific covariates $\mathbf{X}_i$ are considered, a hybrid of EM and Newton-Raphson algorithm suggested by [17] can be used to obtain the numerical ML estimates of regression coefficients. Let $\hat{\boldsymbol{\beta}}_t$ be the value of $\boldsymbol{\beta}$ at the t-th iterations and $S(\hat{\boldsymbol{\beta}}_t)$ and $H(\hat{\boldsymbol{\beta}}_t)$ be the first and second derivative of observed data likelihood with respect to $\boldsymbol{\beta}$ that are evaluated at $[\boldsymbol{\beta}_t, \boldsymbol{\delta}_t, \boldsymbol{\phi}_t, \boldsymbol{\rho}_t]$. Then we can implement Newton-Raphson method as follows:

$$(10)\qquad \hat{\boldsymbol{\beta}}_{t+1} = \hat{\boldsymbol{\beta}}_t - H(\hat{\boldsymbol{\beta}}_t)^{-1} S(\hat{\boldsymbol{\beta}}_t).$$

It is known that the maximum likelihood estimator (MLE) follows asymptotically normal distribution under some regularity conditions. These conditions includes (i) existence of third derivative of log-likelihood function, (ii) boundness of the third derivative of log-likelihood, and (iii) having the Fisher information matrix to be positive-definite. Details of conditions are discussed in [20]. In this paper, we apply suggested conditions in [20] directly and obtain the asymptotic properties of MLE as shown in Lemma 3.1.

Lemma 3.1. *Let $\boldsymbol{\Theta} = [\boldsymbol{\rho}, \boldsymbol{\phi}, \boldsymbol{\beta}, \boldsymbol{\delta}]$ be a vector of free parameters of the LCSM, and let $f(\mathbf{y} \mid \boldsymbol{\Theta})$ be a probability density function of the LCSM. Next, let $\mathbf{y} = [y_1, \ldots, y_n]$ be independent observations, and the log-likelihood equations are give by $l(\boldsymbol{\Theta} \mid \mathbf{y}) = \sum_{i=1}^{n} log f(\mathbf{y}_i \mid \boldsymbol{\Theta})$. Finally, let Θ_0 be the true value of parameters, and let $\hat{\boldsymbol{\Theta}}_n$ be the MLE of $\boldsymbol{\Theta}$. Then, $\hat{\boldsymbol{\Theta}}_n$ is asymptotically normally distributed with mean $\boldsymbol{\Theta}_0$ and covariance matrix $\{nI_n(\hat{\boldsymbol{\Theta}}_n)\}^{-1}$, where $I_n(\hat{\boldsymbol{\Theta}}_n) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial l(\Theta|\mathbf{y}_i)}{\partial \Theta} \frac{\partial l(\Theta|\mathbf{y}_i)}{\partial \Theta}^T \Big|_{\hat{\boldsymbol{\Theta}}_n}$ is the observed Fisher information matrix evaluated at $\hat{\boldsymbol{\Theta}}_n$.*

Lemma 3.1 can be proved by directly applying Proposition 1 in [20]. The observed Fisher information requires the second derivatives of the log-likelihood function with respect to all elements of $\boldsymbol{\Theta}$. The details of the first and second derivatives of log-likelihood functions with respect to $\boldsymbol{\Theta}$ are given in an Appendix.

Bayesian estimation via MCMC algorithm

There are several difficulties in implementing the EM algorithm in a multi-dimensional finite mixture model. First, the EM algorithm may fail to provide a global maximum if the algorithm starts with inappropriate initial values. Second, the Hessian matrix of the model can, numerically, be singular when the parameter estimates locate near the boundary of the parameter space, or the likelihood function does not have a quadratic shape. In addition, the asymptotic properties of ML estimates may not apply if the sample is not sufficiently large. As an alternative approach, we suggest a Bayesian framework which not only may address the obstacles of the EM method, but also can use subjective information on the parameters of the model.

Let $\boldsymbol{\Theta} = [\boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{\phi}, \boldsymbol{\rho}]$ be the vector of parameters of LCSM. The posterior distribution $P(\boldsymbol{\Theta} \mid \mathbf{R}, \mathbf{Y}, \mathbf{X})$ is specified based on the prior distribution $P(\boldsymbol{\Theta})$ and the likelihood function $L(\boldsymbol{\Theta} \mid \mathbf{Y}, \mathbf{R}, \mathbf{X})$. In case of LCSM, however, the posterior distribution is difficult to describe because of unobservable latent variables $[U, W]$. Instead, we may use augmented posterior distribution as suggested in [21, 22]. Consider an augmented likelihood $P(\boldsymbol{\Theta} \mid \mathbf{R}, \mathbf{Y}, \mathbf{Z})$, where $\mathbf{Z}_i = [U_i, W_i]$ is a vector of the latent memberships of the ith individual, $i = 1, \ldots, N$. Such augmented likelihood can be obtained by imputing latent class membership $\mathbf{Z}_i$ based on parameters $\boldsymbol{\Theta}$ and $[\mathbf{Y}_i, \mathbf{R}_i]$. Then we can obtain an augmented posterior probability distribution $P(\boldsymbol{\Theta} \mid \mathbf{R}, \mathbf{Y}, \mathbf{Z})$ by multiplying prior distributions $P(\boldsymbol{\Theta})$ with the augmented likelihood which can be calculated based on the conditional probabilities in Eq. (6).

Based on this augmented posterior distribution, we suggest an iterative two-step MCMC procedure which is a type of Gibbs sampling using the data augmentation. It consists of an imputation step (I-step) and a posterior step (P-step). In the I-step, we calculate the augmented posterior by generating a random $\mathbf{Z}_i$ for $i = 1, \ldots, N$ independently using the current parameters and observed data $[\mathbf{Y}, \mathbf{R}]$. In the P-step, we draw new parameter values from the augmented posterior distribution. Iterating between these two steps provides a sequence of posterior samples of parameters that converge to the posterior distribution, and those posterior samples can be used for parameter estimations.

In the I-step, the latent class membership for the $[U, W]$ can be imputed by drawing $Z_{i(u,w)} = [U_i, W_i]$ independently from a multinomial distribution with the probabilities $\theta_{i(1,1)}, \ldots, \theta_{i(D,S)}$ as follows.

$$(11)\quad [z_{i(1,u)}, \ldots, z_{i(D,u)}] \overset{iid}{\sim} multi(1, \theta_{i(1,u)}, \ldots, \theta_{i(D,u)}), \quad u = 1, \ldots, S.$$

Given the imputed class and profile membership, the augmented likelihood can be written as follows:

$$P(\boldsymbol{\Theta}_{(u,w)} \mid \mathbf{R}, \mathbf{Y}, \mathbf{Z}) \propto \prod_{i=1}^{N} \gamma_d^{I(U_i=u)} \delta_{d|u}^{I(U_i=u, W_i=w)} \times \prod_{i=1}^{N} \prod_{p=1}^{P} \prod_{h=0}^{1} \phi_{ph|w,u}^{I(R_{ip}=h, U_i=u, W_i=w)} \times \prod_{i=1}^{N} \prod_{m \in O_i} \prod_{k=1}^{r_m} \rho_{mk|u}^{I(Y_{im}=k, U_i=u)} \tag{12}$$

In this paper, we impose the Jeffrey's prior on each of the parameters. Let $P(\boldsymbol{\rho}), P(\boldsymbol{\phi}), P(\boldsymbol{\gamma}), P(\boldsymbol{\delta})$ be the prior distributions. Since all random variables $(Y_j \mid U)$, $(R_j \mid W, U)$, $(W \mid U)$, and U follow multinomial distributions, Jeffrey's prior for each set of parameters become Dirichlet distribution as follows

$$\begin{aligned} [\gamma_1, \ldots, \gamma_S] &\sim D(1/2, \ldots, 1/2), \\ [\delta_{1|u}, \ldots, \delta_{D|u}] &\sim D(1/2, \ldots, 1/2), \quad u = 1, \ldots, S. \\ [\rho_{p1|u}, \ldots, \rho_{pr_p|u}] &\sim D(1/2, \ldots, 1/2), \quad p = 1, \ldots, P. \\ [\phi_{p0|w,u}, \phi_{p1|w,u}] &\sim D(1/2, 1/2), \quad p = 1, \ldots, P. \end{aligned} \tag{13}$$

Now, we may calculate posterior distributions as follows:

$$P(\boldsymbol{\Theta}_{(u,w)} \mid \mathbf{R}, \mathbf{Y}, \mathbf{Z}) \propto \delta_{w|u}^{(n_{w|u}+1/2)} \gamma_u^{(n_u+1/2)} \times \prod_{p=1}^{P} \prod_{h=0}^{1} \phi_{ph|w,u}^{(n_{ph|w,u}+1/2)} \times \prod_{m=1}^{P} \prod_{k=1}^{r_m} \rho_{mk|u}^{(n_{mk|u}+1/2)}. \tag{14}$$

Here, the number of observations imputed to the latent classes are defined as follows.

$$\begin{aligned} n_{w|u} &= \sum_{i=1}^{N} I(U_i = u, W_i = w), \quad n_u = \sum_{i=1}^{N} I(U_i = u), \\ n_{mk|u} &= \sum_{i=1}^{N} I(Y_{im} = k, U_i = u) \\ n_{ph|w,u} &= \sum_{i=1}^{N} I(R_{ip} = h, W_i = w, U_i = u) \\ & w = 1, \ldots, D, \quad u = 1, \ldots, S, \quad m, \ p \ = 1, \ldots, P. \end{aligned} \tag{15}$$

In the P-step, we draw new parameter values from Eq. (14) using the Gibbs sampling.

$$\begin{aligned} [\gamma_1, \ldots, \gamma_S] &\sim D(n_1 + 1/2, \ldots, n_S + 1/2), \\ [\delta_{1|u}, \ldots, \delta_{D|u}] &\sim D(n_{(1|u)} + 1/2, \ldots, n_{(D|u)} + 1/2), \\ [\rho_{p1|u}, \ldots, \rho_{pr_p|u}] &\sim D(n_{(p1|u)} + 1/2, \ldots, n_{(pr_p|u)} + 1/2), \\ [\phi_{p0|w,u}, \phi_{p1|w,u}] &\sim D(n_{(p0|w,u)} + 1/2, \ n_{(p1|w,u)} + 1/2). \end{aligned} \tag{16}$$

When we consider the covariate(s) $\mathbf{X}_i$ effect on U, the parameter γ_u is replaced with $\gamma_u(\mathbf{X}_i)$ as suggested in Eq. (5), and $\boldsymbol{\beta}$ becomes a target parameter. Unfortunately, an explicit form of conditional posterior distribution of $P(\boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\rho}, \boldsymbol{\phi} \mid \mathbf{X}_i, \mathbf{Y}_i, \mathbf{R}_i, \mathbf{Z}_i)$ does not exist. Thus, we employ a Metropolis-Hastings algorithm to obtain the posterior sample of $\boldsymbol{\beta}$ instead of using Gibbs sampling. Let $\boldsymbol{\beta}^*$ be a vector of candidate values of $\boldsymbol{\beta}$ from a proposal density $\pi(\boldsymbol{\beta})$, where $\pi(\boldsymbol{\beta})$ is the density function of $N(\hat{\boldsymbol{\beta}}^{ML}, I(\hat{\boldsymbol{\beta}}^{ML})^{-1})$ where $I(\boldsymbol{\beta})$ is the Fisher information of $\boldsymbol{\beta}$. Then, we can calculate the acceptance probability $\alpha(\boldsymbol{\beta}, \boldsymbol{\beta}^*)$ as follows

$$\alpha(\boldsymbol{\beta}, \boldsymbol{\beta}^*) = min(1, \prod_{i=1}^{N} \prod_{u=1}^{S} \left\{ \frac{\gamma_u(\mathbf{X}_i) \mid \boldsymbol{\beta}^*}{\gamma_u(\mathbf{X}_i) \mid \boldsymbol{\beta}} \right\}). \tag{17}$$

The candidate value $\boldsymbol{\beta}^*$ is accepted with probability of Eq. (17) as the iteration proceeds.

Now, we can summarize the Bayesian estimation via Gibbs sampling as follows:

1. Let $\boldsymbol{\Theta}_t = [\boldsymbol{\delta}_t, \boldsymbol{\gamma}_t, \boldsymbol{\phi}_t, \boldsymbol{\rho}_t]$ be parameter values at the t-th iteration.
2. Given observed $\mathbf{Y}_i$ and $\mathbf{X}_i$, calculate conditional probabilities $\theta_{i(u,d)}$ for $i = 1 \ldots N$ and impute the latent class memberships U_i and non-response propensities W_i as in Eq. (6).
3. Based on the imputed latent class memberships $Z_i = [U_i, W_i]$, calculate the augmented likelihood function as in Eq. (12).
4. Calculate the posterior distribution of $\boldsymbol{\Theta}_t$ using the augmented likelihood and prior distributions as in Eq. (14).
5. Draw a new posterior sample $\boldsymbol{\Theta}_{t+1}$ from Eq. (16).
6. Repeat 2. $\sim$ 5. and collect the posterior samples.

To remove the dependency of final parameter estimates on the initial values, we implement sufficiently many MCMC iterations and discard a reasonable number of posterior samples from the beginning as a burn-in period. The length of burn-in period can be determined based on the time-series plot of posterior samples obtained during the iterations.

Numerical Studies

We present two different simulation studies. One is designed to evaluate whether the two estimation strategies for LCSM model work properly. In a single run, we generate synthetic data under LCSM model and fit the model as proposed in Section 3. The 95% credible interval (CI) of each Bayesian estimates is obtained from the posterior samples, while the confidence interval based on the EM method is obtained from the inverse of the negative Hessian matrix. This procedure is repeated 1000 times and the empirical coverage of CIs are obtained. For the true model of the simulation data, we consider a latent class selection model with

Table 1. Simulation results under MNAR, strong parameters, and $N = 1000$.

	EM estimates				MCMC estimates			
Param	Bias	Length	RMSE	CP	Bias	Length	RMSE	CP
$\phi_{11\|11}$	0.033	0.127	0.034	0.947	0.100	0.120	0.031	0.944
$\phi_{21\|11}$	-0.027	0.133	0.032	0.962	0.127	0.128	0.033	0.953
$\phi_{31\|11}$	0.008	0.115	0.031	0.959	-0.113	0.110	0.030	0.935
$\phi_{41\|11}$	0.018	0.117	0.029	0.957	-0.162	0.112	0.029	0.926
$\phi_{11\|21}$	-0.153	0.077	0.021	0.961	0.068	0.069	0.019	0.953
$\phi_{21\|21}$	-0.004	0.070	0.018	0.951	0.098	0.069	0.018	0.944
$\phi_{31\|21}$	0.018	0.070	0.017	0.949	0.011	0.075	0.018	0.953
$\phi_{41\|21}$	-0.057	0.076	0.020	0.949	0.011	0.068	0.017	0.953
$\phi_{11\|12}$	0.116	0.071	0.018	0.947	0.092	0.069	0.016	0.925
$\phi_{21\|12}$	0.024	0.070	0.021	0.942	0.092	0.073	0.018	0.953
$\phi_{31\|12}$	-0.145	0.077	0.018	0.943	0.145	0.069	0.022	0.913
$\phi_{41\|12}$	0.109	0.076	0.030	0.957	0.183	0.068	0.020	0.953
$\phi_{11\|22}$	-0.154	0.117	0.036	0.954	-0.232	0.115	0.029	0.935
$\phi_{21\|22}$	-0.126	0.136	0.029	0.945	-0.002	0.129	0.029	0.981
$\phi_{31\|22}$	0.032	0.125	0.029	0.951	0.269	0.123	0.034	0.953
$\phi_{41\|22}$	0.141	0.118	0.030	0.966	-0.218	0.118	0.031	0.953
$\rho_{11\|1}$	0.100	0.105	0.028	0.961	-0.098	0.104	0.025	0.935
$\rho_{21\|1}$	0.031	0.104	0.026	0.949	-0.011	0.101	0.027	0.935
$\rho_{31\|1}$	0.094	0.073	0.016	0.957	-0.324	0.072	0.021	0.928
$\rho_{41\|1}$	0.118	0.073	0.019	0.956	0.037	0.070	0.017	0.963
$\rho_{11\|2}$	-0.046	0.072	0.019	0.952	-0.200	0.071	0.018	0.965
$\rho_{21\|2}$	0.063	0.105	0.025	0.949	-0.065	0.101	0.028	0.953
$\rho_{31\|2}$	0.093	0.105	0.025	0.941	0.047	0.101	0.026	0.949
$\rho_{41\|2}$	0.172	0.072	0.019	0.958	0.069	0.072	0.018	0.941
$\delta_{1\|1}$	0.033	0.087	0.023	0.962	0.160	0.070	0.023	0.944
$\delta_{1\|2}$	-0.045	0.088	0.022	0.941	0.116	0.091	0.021	0.981
β_{01}	-0.022	0.464	0.256	0.935	-0.183	0.462	0.125	0.925
β_{11}	0.078	0.352	0.170	0.944	0.178	0.346	0.102	0.917

four binary response variables $[Y_1 \dots Y_4]$ and a categorical latent variable U that has two classes ($S = 2$). The size of each simulated data is either 300 or 1000, to represent a small sample size ($N = 300$) and a large sample ($N = 1000$). Finally, the nonresponse propensity W is set to be a binary latent variable (i.e., $D = 2$).

In each scenario, three classes of parameter values are chosen where true ρ-parameter values are (i) strong ($\rho =$ 0.9 or 0.1), (ii) mixed (some values are close to 0.5), and (iii) weak (all values are close to 0.5). For each scenario, we report standardized bias (Bias), root mean-square error (RMSE), coverage (CP), and average length of the interval 95% CI (Length). We consider any standardized bias with an absolute value greater than 0.4 and coverage probabilities that are less than 0.9 to be unacceptable, as suggested in [23]. The proportion of incomplete individuals are set to be 10%, 25%, and 50%. For brevity, we report the results from 50% missing, which is the largest amount of missing values among our scenarios.

Tables 1 through 4 illustrate the simulation results under the strong and mixed pararmeter classes with $N = 300$ and 1000 under the MNAR scenario. Simulation results are acceptable in that the Bias of the estimates are mostly less

Table 2. Simulation results under MNAR, mixed parameters, and $N = 1000$.

	EM estimates				MCMC estimates			
Param	Bias	Length	RMSE	CP	Bias	Length	RMSE	CP
$\phi_{11\|11}$	-0.028	0.181	0.050	0.937	0.019	0.126	0.031	0.933
$\phi_{21\|11}$	-0.164	0.137	0.036	0.942	-0.011	0.167	0.039	0.942
$\phi_{31\|11}$	0.042	0.127	0.031	0.939	-0.125	0.168	0.031	0.942
$\phi_{41\|11}$	0.005	0.124	0.033	0.944	-0.020	0.132	0.031	0.956
$\phi_{11\|21}$	-0.059	0.078	0.019	0.942	-0.022	0.119	0.019	0.946
$\phi_{21\|21}$	-0.093	0.073	0.018	0.956	-0.040	0.116	0.019	0.956
$\phi_{31\|21}$	-0.032	0.075	0.019	0.947	0.107	0.078	0.017	0.956
$\phi_{41\|21}$	0.095	0.078	0.020	0.955	-0.014	0.071	0.019	0.937
$\phi_{11\|12}$	-0.112	0.073	0.020	0.954	-0.096	0.072	0.020	0.966
$\phi_{21\|12}$	0.151	0.075	0.018	0.954	-0.008	0.077	0.017	0.961
$\phi_{31\|12}$	-0.213	0.079	0.021	0.939	-0.026	0.072	0.019	0.963
$\phi_{41\|12}$	0.089	0.110	0.027	0.952	0.097	0.073	0.020	0.943
$\phi_{11\|22}$	-0.035	0.125	0.035	0.962	-0.316	0.078	0.029	0.928
$\phi_{21\|22}$	-0.039	0.239	0.057	0.947	-0.085	0.103	0.034	0.946
$\phi_{31\|22}$	-0.034	0.179	0.057	0.956	0.026	0.121	0.055	0.965
$\phi_{41\|22}$	-0.092	0.133	0.049	0.932	-0.142	0.191	0.044	0.932
$\rho_{11\|1}$	-0.105	0.123	0.033	0.953	-0.010	0.114	0.031	0.956
$\rho_{21\|1}$	0.026	0.152	0.042	0.959	0.109	0.149	0.042	0.932
$\rho_{31\|1}$	0.018	0.108	0.027	0.958	-0.116	0.104	0.027	0.956
$\rho_{41\|1}$	0.047	0.080	0.022	0.948	0.004	0.079	0.019	0.942
$\rho_{11\|2}$	-0.232	0.075	0.021	0.941	0.026	0.074	0.021	0.927
$\rho_{21\|2}$	0.019	0.153	0.041	0.939	-0.105	0.152	0.040	0.937
$\rho_{31\|2}$	0.045	0.157	0.041	0.952	0.062	0.153	0.020	0.923
$\rho_{41\|2}$	0.040	0.076	0.018	0.954	-0.074	0.074	0.027	0.924
$\delta_{1\|1}$	0.151	0.096	0.026	0.957	-0.164	0.101	0.029	0.922
$\delta_{1\|2}$	0.136	0.103	0.027	0.962	0.019	0.106	0.029	0.928
β_{01}	0.052	0.504	0.270	0.934	0.061	0.488	0.145	0.907
β_{11}	-0.001	0.369	0.192	0.935	-0.083	0.363	0.112	0.916

Table 3. Simulation results under MNAR, strong parameters, and $N = 300$.

	EM estimates				MCMC estimates			
Param	Bias	Length	RMSE	CP	Bias	Length	RMSE	CP
$\phi_{11\|11}$	0.045	0.289	0.066	0.981	0.139	0.228	0.054	0.971
$\phi_{21\|11}$	0.181	0.298	0.062	0.906	0.104	0.229	0.058	0.955
$\phi_{31\|11}$	-0.057	0.284	0.070	0.925	-0.296	0.223	0.064	0.919
$\phi_{41\|11}$	-0.144	0.265	0.067	0.887	-0.226	0.225	0.057	0.923
$\phi_{11\|21}$	0.084	0.159	0.041	0.878	0.002	0.147	0.042	0.927
$\phi_{21\|21}$	0.054	0.152	0.040	0.934	-0.199	0.138	0.033	0.971
$\phi_{31\|21}$	0.006	0.153	0.034	0.943	-0.178	0.137	0.034	0.961
$\phi_{41\|21}$	0.035	0.157	0.040	0.953	0.015	0.142	0.041	0.953
$\phi_{11\|12}$	-0.038	0.149	0.039	0.916	-0.095	0.139	0.040	0.934
$\phi_{21\|12}$	-0.063	0.144	0.032	0.925	-0.167	0.138	0.040	0.943
$\phi_{31\|12}$	-0.206	0.160	0.041	0.934	0.268	0.152	0.039	0.929
$\phi_{41\|12}$	-0.009	0.161	0.037	0.935	0.048	0.142	0.045	0.924
$\phi_{11\|22}$	-0.018	0.292	0.062	0.972	-0.209	0.212	0.039	0.953
$\phi_{21\|22}$	-0.024	0.301	0.064	0.915	0.168	0.228	0.056	0.957
$\phi_{31\|22}$	-0.009	0.282	0.058	0.962	0.198	0.226	0.062	0.962
$\phi_{41\|22}$	-0.007	0.291	0.06	0.962	-0.217	0.212	0.053	0.944
$\rho_{11\|1}$	-0.032	0.232	0.049	0.961	0.252	0.141	0.054	0.934
$\rho_{21\|1}$	0.051	0.224	0.051	0.962	0.253	0.147	0.056	0.943
$\rho_{31\|1}$	0.126	0.152	0.035	0.925	-0.187	0.138	0.035	0.949
$\rho_{41\|1}$	0.208	0.155	0.037	0.962	-0.313	0.196	0.041	0.949
$\rho_{11\|2}$	0.153	0.154	0.033	0.953	-0.135	0.199	0.037	0.962
$\rho_{21\|2}$	0.098	0.235	0.050	0.972	-0.257	0.138	0.056	0.916
$\rho_{31\|2}$	0.186	0.223	0.061	0.943	0.249	0.180	0.055	0.928
$\rho_{41\|2}$	0.052	0.148	0.033	0.887	0.084	0.179	0.038	0.937
$\delta_{1\|1}$	0.042	0.180	0.049	0.962	0.091	0.201	0.048	0.953
$\delta_{1\|2}$	0.022	0.182	0.041	0.953	-0.221	0.198	0.046	0.959
β_{01}	-0.193	0.945	0.256	0.935	-0.014	0.934	0.253	0.917
β_{11}	0.231	0.714	0.170	0.944	-0.025	0.709	0.214	0.907

Table 4. Simulation results under MNAR, mixed parameters, and $N = 300$.

	EM estimates				MCMC estimates			
Param	Bias	Length	RMSE	CP	Bias	Length	RMSE	CP
$\phi_{11\|11}$	0.024	0.260	0.075	0.943	0.249	0.261	0.063	0.962
$\phi_{21\|11}$	0.037	0.239	0.064	0.935	-0.136	0.244	0.057	0.971
$\phi_{31\|11}$	-0.149	0.281	0.069	**0.896**	-0.279	0.224	0.058	0.972
$\phi_{41\|11}$	-0.001	0.294	0.057	0.953	0.210	0.224	0.054	0.934
$\phi_{11\|21}$	-0.061	0.163	0.039	0.942	-0.119	0.154	0.039	0.943
$\phi_{21\|21}$	0.033	0.155	0.039	0.926	-0.049	0.139	0.038	0.962
$\phi_{31\|21}$	0.015	0.153	0.038	0.943	0.184	0.141	0.040	0.941
$\phi_{41\|21}$	-0.276	0.169	0.036	0.981	-0.050	0.150	0.039	0.932
$\phi_{11\|12}$	0.091	0.154	0.033	0.972	-0.289	0.139	0.039	0.934
$\phi_{21\|12}$	0.057	0.157	0.042	0.915	-0.104	0.147	0.037	0.980
$\phi_{31\|12}$	-0.080	0.162	0.037	0.943	-0.006	0.146	0.039	0.966
$\phi_{41\|12}$	-0.149	0.210	0.053	0.981	-0.208	0.165	0.041	0.932
$\phi_{11\|22}$	0.154	0.311	0.039	0.925	0.229	0.225	0.063	0.934
$\phi_{21\|22}$	0.039	0.444	0.038	0.962	0.272	0.281	0.060	0.990
$\phi_{31\|22}$	-0.126	0.381	0.036	0.971	-0.250	0.261	0.062	0.931
$\phi_{41\|22}$	0.026	0.303	0.033	0.925	0.292	0.238	0.061	0.935
$\rho_{11\|1}$	0.180	0.257	0.042	0.925	0.207	0.227	0.068	0.916
$\rho_{21\|1}$	-0.015	0.311	0.037	0.972	0.106	0.286	0.076	0.943
$\rho_{31\|1}$	-0.039	0.213	0.053	0.972	-0.181	0.205	0.057	0.916
$\rho_{41\|1}$	0.095	0.165	0.068	0.981	-0.127	0.151	0.042	0.912
$\rho_{11\|2}$	-0.011	0.156	0.092	0.953	-0.224	0.146	0.039	0.925
$\rho_{21\|2}$	0.022	0.308	0.075	0.943	-0.109	0.294	0.079	0.927
$\rho_{31\|2}$	-0.062	0.319	0.067	0.990	0.168	0.302	0.076	0.919
$\rho_{41\|2}$	-0.032	0.156	0.036	0.953	0.026	0.143	0.040	0.981
$\delta_{1\|1}$	0.054	0.194	0.047	0.969	-0.155	0.189	0.047	0.952
$\delta_{1\|2}$	-0.047	0.202	0.057	**0.897**	-0.292	0.194	0.051	0.971
β_{01}	-0.158	1.050	0.270	0.934	-0.124	1.023	0.270	0.934
β_{11}	0.209	0.754	0.192	0.935	0.195	0.767	0.218	0.916

Table 5. Simulation results using MCMC algorithm under MNAR, weak parameters.

	$N = 1000$				$N = 300$			
Param	Bias	Length	RMSE	CP	Bias	Length	RMSE	CP
$\phi_{11\|11}$	0.068	0.175	0.044	0.942	0.337	0.242	0.060	0.963
$\phi_{21\|11}$	0.024	0.132	0.038	0.928	-0.346	0.280	0.078	0.924
$\phi_{31\|11}$	-0.003	0.121	0.033	0.931	0.062	0.245	0.060	0.946
$\phi_{41\|11}$	-0.004	0.118	0.031	0.928	0.321	0.273	0.081	0.943
$\phi_{11\|21}$	0.014	0.079	0.022	0.928	-0.245	0.182	0.050	0.925
$\phi_{21\|21}$	-0.002	0.071	0.018	0.952	-0.330	0.161	0.040	0.963
$\phi_{31\|21}$	-0.002	0.073	0.019	0.956	-0.001	0.162	0.047	0.944
$\phi_{41\|21}$	-0.002	0.077	0.021	0.924	-0.040	0.170	0.045	0.921
$\phi_{11\|12}$	0.016	0.077	0.017	0.966	-0.002	0.165	0.047	0.925
$\phi_{21\|12}$	-0.003	0.072	0.018	0.953	0.042	0.154	0.044	0.927
$\phi_{31\|12}$	-0.003	0.074	0.019	0.961	0.159	0.184	0.048	0.944
$\phi_{41\|12}$	0.017	0.078	0.021	0.951	-0.183	0.162	0.039	0.925
$\phi_{11\|22}$	-0.014	0.104	0.029	0.961	-0.265	0.249	0.060	0.953
$\phi_{21\|22}$	-0.001	0.120	0.030	0.927	0.252	0.279	0.069	0.971
$\phi_{31\|22}$	0.012	0.170	0.046	0.946	0.266	0.252	0.068	0.970
$\phi_{41\|22}$	-0.004	0.126	0.029	0.942	-0.151	0.248	0.065	0.962
$\rho_{11\|1}$	0.012	0.120	0.034	0.923	0.180	0.315	0.081	0.953
$\rho_{21\|1}$	0.010	0.149	0.038	0.956	0.017	0.308	0.088	0.916
$\rho_{31\|1}$	-0.001	0.105	0.031	0.949	-0.114	0.215	0.048	0.972
$\rho_{41\|1}$	-0.002	0.079	0.021	0.932	-0.084	0.215	0.057	0.933
$\rho_{11\|2}$	-0.003	0.076	0.020	0.946	-0.206	0.219	0.058	0.971
$\rho_{21\|2}$	-0.003	0.153	0.042	0.937	-0.012	0.309	0.078	0.944
$\rho_{31\|2}$	0.013	0.154	0.039	0.946	0.044	0.312	0.084	0.961
$\rho_{41\|2}$	0.025	0.075	0.020	0.942	0.236	0.221	0.060	0.962
$\delta_{1\|1}$	0.091	0.201	0.048	0.953	0.307	0.209	0.048	0.972
$\delta_{1\|2}$	-0.221	0.198	0.046	0.959	-0.245	0.210	0.058	0.925
β_{01}	0.065	0.934	0.253	0.917	0.065	1.176	0.292	0.944
β_{11}	0.026	0.709	0.214	0.907	0.026	0.812	0.117	0.971

Table 6. The standardized bias of four different methods with weak measurement parameterers.

	$N = 1000$				$N = 300$			
Param	EM	MCMC	MAR	CCA	EM	MCMC	MAR	CCA
β_{10}	-0.095	-0.051	-0.067	-0.058	-0.156	-0.154	-0.036	-0.058
β_{11}	0.166	0.149	0.216	0.305	0.200	0.065	0.166	0.196
$\phi_{11\|1}$	-0.081	0.041	-0.053	-0.018	-0.086	-0.059	-0.100	0.264
$\phi_{21\|1}$	-0.019	-0.009	-0.005	0.077	-0.196	-0.130	-0.156	0.238
$\phi_{31\|1}$	0.020	0.021	-0.019	-0.081	0.008	0.023	-0.001	**-0.418**
$\phi_{41\|1}$	0.128	0.073	0.098	0.117	0.052	0.062	-0.036	**-0.416**
$\phi_{11\|2}$	-0.017	0.011	-0.020	-0.018	0.057	0.056	0.109	-0.252
$\phi_{21\|2}$	0.029	0.040	0.026	0.032	0.088	0.065	0.091	-0.242
$\phi_{31\|2}$	-0.059	-0.042	-0.088	-0.111	-0.033	-0.017	-0.021	0.314
$\phi_{41\|2}$	-0.139	-0.096	-0.153	-0.160	-0.040	-0.049	-0.044	0.244

than 0.4, and CP are close to 0.95 for all parameters in both small and large sample size. Such results are expected because the synthetic data are generated from the fitted model. These tables imply that both EM and MCMC algorithms provide proper point estimates and confidence intervals under strong and mixed scenarios. When the sample size is small (i.e., $N = 300$), the standardized bias tend to increase and coverage probabilities are more deviated from 0.95 than when $N = 1000$, but these results are still acceptable if the parameters are still in the strong class. In the case of the mixed strength parameter class with $N = 300$, the coverage probabilities of some ML estimates from the EM algorithm are lower than 0.9, as highlighted in Table 4. On the other hand, the performance of Bayesian estimates from MCMC algorithm are acceptable for all of the classes of parameters.

In the strong and mixed scenarios, the Bias and RMSE of the two estimation methods are similar. The Bayesian estimates shows slightly shorter CIs than ML estimates, but the differences are hardly noticeable. In the weak scenarios, however, the Hessian matrix of LCSM are mostly singular when evaluated at the MLEs and thus the asymptotic confidence intervals for ML estimates are unavailable. Meanwhile, the credible intervals for the Bayesian estimates showed reasonable coverage probabilities. Table 5 conveys the simulation results of MCMC methods under the weak scenario with $N = 300$ and 1000. Since the biases are reasonably small and the coverage probabilities are all close to 0.95, we can conclude that MCMC algorithm works properly, regardless of the sample size and the class of parameters. In this sense, the Bayesian estimation is preferred to EM methods especially when the measurement parameters are weak.

A second simulation study is conducted to compare the performance of several methods with different assumptions on the missingness mechanisms. We generate the data using the same latent class structure as in the first study, and fitted three different methods as follows: (1) the proposed latent class selection model using EM-algorithm (EM), (2) the proposed latent class selection model using MCMC-procedure (MCMC), (3) a latent class model with adjusted

Table 7. The average length of interval of four different methods with weak measurement parameterers.

	$N = 1000$				$N = 300$			
Param	EM	MCMC	MAR	CCA	EM	MCMC	MAR	CCA
β_{10}	1.451	1.126	1.304	1.510	2.212	2.056	2.631	3.704
β_{11}	0.849	0.632	0.677	0.825	2.371	1.378	1.433	2.760
$\phi_{11\|1}$	0.174	0.146	0.155	0.178	0.342	0.256	0.300	0.379
$\phi_{21\|1}$	0.175	0.147	0.156	0.178	0.347	0.256	0.296	0.384
$\phi_{31\|1}$	0.175	0.147	0.156	0.178	0.342	0.250	0.289	0.388
$\phi_{41\|1}$	0.178	0.148	0.156	0.178	0.340	0.256	0.293	0.380
$\phi_{11\|2}$	0.178	0.148	0.158	0.181	0.344	0.251	0.291	0.381
$\phi_{21\|2}$	0.177	0.146	0.157	0.180	0.346	0.254	0.292	0.381
$\phi_{31\|2}$	0.180	0.148	0.157	0.181	0.345	0.256	0.290	0.383
$\phi_{41\|2}$	0.179	0.148	0.157	0.180	0.345	0.253	0.290	0.385

Table 8. The RMSE of four different methods with weak measurement parameterers.

	$N = 1000$				$N = 300$			
Param	EM	MCMC	MAR	CCA	EM	MCMC	MAR	CCA
β_{10}	0.553	0.369	0.367	0.424	1.049	0.841	0.831	1.293
β_{11}	0.252	0.188	0.183	0.230	0.578	0.407	0.414	1.266
$\phi_{11\|1}$	0.054	0.045	0.043	0.048	0.127	0.083	0.079	0.095
$\phi_{21\|1}$	0.048	0.042	0.041	0.042	0.134	0.086	0.081	0.097
$\phi_{31\|1}$	0.052	0.043	0.042	0.046	0.126	0.079	0.072	0.091
$\phi_{41\|1}$	0.054	0.041	0.041	0.047	0.125	0.084	0.077	0.084
$\phi_{11\|2}$	0.059	0.044	0.044	0.051	0.133	0.083	0.079	0.097
$\phi_{21\|2}$	0.051	0.041	0.042	0.045	0.136	0.086	0.085	0.098
$\phi_{31\|2}$	0.054	0.042	0.040	0.046	0.134	0.084	0.079	0.088
$\phi_{41\|2}$	0.053	0.043	0.042	0.050	0.133	0.081	0.076	0.091

Table 9. The coverage probability of intervals of four different methods with weak measurement parameterers.

	$N = 1000$				$N = 300$			
Param	EM	MCMC	MAR	CCA	EM	MCMC	MAR	CCA
β_{10}	**0.892**	0.929	0.919	**0.874**	**0.879**	0.922	0.921	0.940
β_{11}	**0.844**	0.932	0.959	0.954	**0.890**	0.974	0.928	0.905
$\phi_{11\|1}$	**0.893**	0.941	0.929	0.929	**0.874**	0.973	0.917	0.883
$\phi_{21\|1}$	0.914	0.936	0.964	0.949	**0.862**	0.923	0.926	0.889
$\phi_{31\|1}$	**0.793**	0.959	0.959	0.954	**0.863**	0.921	0.919	0.910
$\phi_{41\|1}$	**0.795**	0.924	0.929	0.957	**0.873**	0.945	0.926	0.935
$\phi_{11\|2}$	**0.893**	0.969	0.944	0.935	**0.847**	0.962	0.913	0.913
$\phi_{21\|2}$	0.929	0.924	0.959	0.954	**0.839**	0.967	0.923	0.905
$\phi_{31\|2}$	**0.894**	0.964	0.954	**0.899**	**0.865**	0.957	0.924	**0.825**
$\phi_{41\|2}$	**0.893**	0.939	0.932	**0.899**	**0.866**	0.959	0.915	**0.891**

estimator under MAR mechanism (MAR), as used in [24], and (4) the conventional latent class model using only the complete cases (CCA).

When the simulation data are generated from the strong/mixed classes of parameters with $N = 1000$, the performance of all four methods are comparable. CCA tends to have larger biases and longer CIs than the other methods, but the magnitude of the differences is barely noticeable. When the simulation is conducted with the weak classes of parameters, however, the results are quite different. Hence, we only report the weak parameter class scenarios for brevity.

Tables 6∼9 display the standardized bias, average interval length, root mean-square error, and coverage probability of 95% CIs from four different methods under weak-measurement parameters, respectively. Overall, CCA provides the largest standardized biases, and some of them are problematic in that their absolute values exceed 0.4 [23]. Also, CCA provides the largest length of intervals among the four methods. Consequently, confidence intervals based on the CCA method fail to cover the true parameter values too often, as shown in Table 9. With the MAR method, biases are acceptable but the lengths of the confidence intervals tend to be inflated compared to the MCMC method, resulting in lower coverage probabilities. Such trends are more noticeable when the sample size is small.

The performance of EM-algorithm on our proposed LCSM are also questionable when the population is generated under the weak parameter class. As mentioned in the first simulation study, the information matrix is singular so the standard errors are unavailable. In the simulation study, we inverted the sub-matrix of Hessian matrix that corresponds to data parameters only (i.e., $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$) to compare the results to that of MAR and CCA. As shown in Table 9, the performance of EM estimates are questionable in that the coverage probabilities noticeably deviate from 0.95 under weak-measurement scenario regardless of the sample size. In addition, average interval lengths, RMSE, and standardized biases are noticeably larger than other three methods. On the other hand, the Bayesian estimates via proposed MCMC algorithm show acceptable coverage probabilities between 0.922 $\sim$ 0.973. Also, their average length of intervals and RMSE are smaller than the others. Consequently, we conclude that using MCMC algorithm would be a better choice when the estimated measurement parameters are weak (that is, between 0.3 and 0.7). Under the weak-measurement parameters, the performance of EM-estimates are unacceptable in terms of standard errors and other four measures. The Bayesian estimates, on the other hand, work properly, as discussed earlier in Table 5.

4. APPLICATION TO GSS DATA

The General Social Study (GSS) is a longitudinal sociological survey that monitors and explains trends, behaviors, and attributes among non-institutionalized adults in

the United States. The GSS has published the updated data annually or biennially since 1972, and the latest version was published in 2018 with questionnaires regarding social issues such as mental health, religions, child stigma, quality of work-life etc. The data sets are available at the GSS website: *http://www.gss.norc.org/*.

Numerous papers studied public opinions and attitudes toward legal abortion based on GSS data [25, 26, 27]. Along with these papers, we apply our proposed LCSM to illustrate its usage in statistical practice by discovering new insights on abortion attitudes among U.S. citizens. In the 2018 GSS data, respondents were asked whether or not it should be legal for a pregnant woman to have an abortion under specific circumstances. Six binary variables are defined as "Tell me whether or not you think it should be possible for a pregnant woman to obtain a legal abortion under certain circumstances?" Six circumstances are (1) for any reasons (*ABANY*), (2) if there is a strong chance of serious defect in the baby (*ABDEFECT*), (3) if the woman's own health is seriously endangered by the pregnancy (*ABHLTH*), (4) if the woman is married and does not want any more children (*ABNOMORE*), (5) if the family has a very low income and cannot afford any more children (*ABPOOR*), and (6) if the woman became pregnant as a result of rape (*ABRAPE*). Table 10 enumerates the proportions of 1,574 individuals who responded "Yes" and nonresponse proportions under six particular circumstances.

Table 10. Proportion of "Yes" to questionnaires: "Should the legal abortion be possible for a pregnant woman to obtain a legal abortion under certain circumstances*?"*

Item	Yes (%)	Missing (%)
ABANY	48.53	3.17
ABDEFECT	73.95	4.51
ABHLTH	49.43	3.81
ABNOMORE	86.78	3.62
ABPOOR	47.65	3.43
ABRAPE	75.86	4.57

In our model, the number of latent classes S for a latent variable U and the number of latent propensity D for W need to be determined when implementing proposed LCSM. We obtain the maximum likelihood estimates of LCSM with different number of classes using the EM algorithm and calculate the Bayesian Information Criterion (BIC). The four-class model is chosen since it has the lowest BIC. Table 11 illustrates the estimated BIC for different number of classes.

Table 11. BIC for latent class selection models with different number of classes

# of class	2	3	4	5	6
BIC	9778.13	9543.40	**9457.58**	9759.47	9905.55

Based on the latent class structure, we fit a latent class selection model using the MCMC algorithm with Jeffrey's priors. Both estimation methods provide similar estimates, but the performance of Bayesian estimates are slightly better in terms of RMSE. As such, we discuss the parameter estimates from the MCMC algorithm. Table 12 conveys the estimated measurement parameters (i.e., $\hat{\rho}_{p1|d}$, $p = 1 \dots 6$, $d = 1 \dots 4$) and prevalence of the four class model. As discussed in Section 2, each estimated $\hat{\rho}_{p1|d}$ denotes the probability of agreement, "Yes", with each questionnaire item under the dth latent class. *Class 1* showed high probabilities of agreement to legal abortions for all 6 items, while *Class 4* strongly disagreed on all of them. In this sense, *Class 1* may be interpreted as *Liberal (pro-choice)*, while *Class 4* can be *Conservative (pro-life)*. Individuals in *Class 2* and *Class 3* showed high probabilities of consent to the legal abortion in *ABDEFECT*, *ABNOMORE*, and *ABRAPE*, but the estimated probabilities were generally higher in *Class 2* than in *Class 3*. In this sense, *Class 2* is interpreted as *Partially supportive*, and *Class 3* as *Weakly supportive*.

Table 12. Parameter estimates of 4-class model for patterns

Manifest item	Latent class for response variables			
	Class 1	Class 2	Class 3	Class 4
ABANY	0.948	0.419	0.027	0.012
ABDEFECT	0.986	0.907	0.637	0.080
ABHLTH	0.980	0.447	0.006	0.005
ABNOMORE	0.997	0.961	0.987	0.245
ABPOOR	0.974	0.322	0.028	0.004
ABRAPE	0.994	0.907	0.687	0.079
Class proportion	0.436	0.203	0.236	0.125

We also considered the effect of three demographic factors (age, gender, and race) on the prevalence of the latent classes. Table 13 conveys the estimated odds ratios for each latent classes versus *Class 1* and their 95% CIs. The odds ratios of uth latent class for ith individual were calculated as $exp(\mathbf{X}\boldsymbol{\beta}_u)$, where $\boldsymbol{\beta}_u$ is multinomial logistic regression coefficients of uth latent class, $u = 2 \dots 4$. Based on the 95% CI, we conclude that the race variable had significant effect on the prevalences of the latent classes. The odds ratio of *Class 2* vs *Class 1* was 1.385 times higher in *Black* respondents than *White respondents*. Similarly, the odds ratio of *Class 4* vs *Class 1* was 1.506 times higher in *Black* respondents than *White respondents*. Lastly, the prevalence of each latent class were obtained based on Eq. (5) using the estimated regression model.

Finally, each of the four latent classes had two subclasses, where one denotes the complete group and the other describes nonresponse group. Table 14 conveys the estimated probabilities of "being observed" of each items ($\phi_{p1|w,u}$, p=1,...,6, u=1,...,4, w = 1, 2) and corresponding prevalence. Prevalences of each sub-class were calculated based on the proportions of latent classes in Table 13 and estimated conditional probabilities (i.e., $\delta_{w|u}$, w=1, 2, u=1,...,4). The estimated $\hat{\phi}$ of observed subclasses are similar across all four latent classes, so we impose the following constraints:

Table 13. Estimated odds ratios and 95% CI for each latent classes

	Class 2	Class 3	Class 4
Proportion	0.203	0.236	0.125
Intercept	0.406 [0.254, 0.649]	0.753 [0.452, 1.253]	0.208 [0.127, 0.339]
Age	0.998 [0.990, 1.006]	0.994 [0.986, 1.003]	1.004 [0.997, 1.012]
Female(vs Male)	1.264 [0.988, 1.619]	0.889 [0.674, 1.175]	1.109 [0.858, 1.435]
Black(vs White)	1.385 [1.023, 1.876]	0.972 [0.665, 1.418]	1.506 [1.117, 2.031]
Others(vs White)	1.132 [0.773, 1.658]	0.974 [0.624, 1.519]	0.752 [0.483, 1.173]

$\phi_{p1|1,1} = \phi_{p1|1,2} = \cdots = \phi_{p1|1,4}$, $p = 1 \ldots 6$. Such constraints provided a simplified model by identifying a single complete-subclass and four incomplete subclasses which explained nonresponse patterns. The likelihood ratio test for $H_0 : \phi_{p1|1,1} = \phi_{p1|1,2} = \cdots = \phi_{p1|1,4}$, $p = 1 \ldots 6$, vs H_a: Not H_0. is not rejected at $\alpha = 0.05$ (test statistic $X^2 = 18.804$, $df = 18$) so we conclude that these constraints are acceptable.

Individuals who belonged to the *Completers* group show high probabilities of response for all six items. The subclass of *Class 1* show very low probabilities (i.e., $0.039 \sim 0.146$) of being observed for each item. Next, individuals in the subclass of *Class 2* are unlikely to respond to *ABHLTH*, *ABPOOR*. Similarly, the subclass of *Class 3* have low probabilities of response to *ABDEFECT*, *ABRAPE*. Finally, individuals in the subclass of *Class 4* are unlikely to respond to *ABMORE* only, with 0.166 probability of response.

It is well known that the finite mixtures of distributions does not satisfy the identifiability condition because permutations of parameter labels yield identical density functions. Instead, researchers have focused on providing a local identifiability of the latent class model. Huang [18] discussed the local identifiability of the conventional latent class model with covariate and constrained parameters. In this paper, we investigate the local identifiability of the fitted model by applying the conditions in [18] directly. To obtain global identifiability, appropriate constraints that incorporate scientific knowledge and theory are needed [18, 28].

Let $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_4]$ be a 115 by 4 dimensional matrix. Here, 115 is the number of unique response patterns from $[Y_1, \ldots, Y_6]$, and $\boldsymbol{\phi}_d \in \mathbb{R}^{115}$ is a vector of probabilities of the 115 unique response patterns under dth latent class, $d = 1, \ldots, 4$. Also, let $\boldsymbol{\psi} = [\boldsymbol{\psi}_{1|1}, \ldots, \boldsymbol{\psi}_{2|4}]$ be a 38 by 8 dimensional matrix. Here, 38 is the number of unique response patterns from $[R_1, \ldots, R_6]$, and $\boldsymbol{\psi}_{f|d} \in \mathbb{R}^{38}$ is a vector of probabilities of the 38 unique response patterns under dth latent class and fth latent missing propensity, $d = 1, \ldots, 4$, $f = 1, 2$. The fitted latent class selection model is locally identifiable if (i) $[\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_4]$ are linearly independent and (ii) $\boldsymbol{\psi} = [\boldsymbol{\psi}_{1|1}, \ldots, \boldsymbol{\psi}_{2|4}]$ are linearly independent. We calculated the eigenvalues of $\boldsymbol{\Phi}^T\boldsymbol{\Phi}$ and $\boldsymbol{\psi}^T\boldsymbol{\psi}$ and observed that all eigenvalues are positive. This means that our fitted model is locally identifiable.

Table 14. Parameter estimates for nonresponse patterns of 4-class model

Nonresponse propensity	Complete	Subclass of Class 1	Subclass of Class 2	Subclass of Class 3	Subclass of Class 4
ABANY	0.992	0.039	0.622	0.750	0.867
ABDEFECT	0.983	0.146	0.650	0.488	0.735
ABHLTH	0.990	0.029	0.346	0.833	0.881
ABNOMORE	0.984	0.099	0.980	0.523	0.166
ABPOOR	0.992	0.094	0.441	0.880	0.772
ABRAPE	0.986	0.151	0.834	0.142	0.501
Proportion	0.929	0.011	0.028	0.021	0.011

As an alternative approach, we fit a latent class model under an MNAR assumption as suggested in [15]. In this model, both questionnaire items and their missing indicators (i.e., $R_{ip} = 1$ if Y_{ip} is observed, for all i, p) are treated as response variables. Table 15 shows the 4-class LCA model that consists of original 6 items and corresponding missingness indicators. Similar to our suggested model in Table 12, *Class 1* was strongly in favor of the legal abortion, while *Class 4* hardly agreed. *Class 2* and *Class 3* had similar response patterns to the questionnaires but *Class 3* was unlikely to respond to all items except *ABNOMORE*. Based on Table 15, we found that the alternative model and our proposed LCSM provided similar interpretations for the four latent classes, but the proposed LCSM identified more detailed nonresponse propensities by discovering four distinct subclasses that showed different missing patterns. On the other hand, the alternative model provided a simpler structures in that it had smaller number of classes and each latent class was defined based on both response patterns and nonresponse patterns simultaneously.

Table 15. Parameter estimates of 4-class model.

Latent Class	Class 1	Class 2	Class 3	Class 4
ABANY	0.923	0.134	0.192	0.008
ABDEFECT	0.984	0.778	0.884	0.094
ABHLTH	0.960	0.113	0.123	0.004
ABNOMORE	0.993	0.975	0.978	0.379
ABPOOR	0.936	0.087	0.240	0.003
ABRAPE	0.990	0.802	0.961	0.097
ABANY.R	0.992	0.987	0.434	0.993
ABDEFECT.R	0.986	0.975	0.370	0.969
ABHLTH.R	0.993	0.980	0.320	0.991
ABNOMORE.R	0.999	0.996	0.589	0.881
ABPOOR.R	0.993	0.984	0.412	0.989
ABRAPE.R	0.997	0.972	0.442	0.919
Proportion	0.477	0.325	0.042	0.156

5. CONCLUSIONS

Latent class models are popular tools for discovering populations' latent groups with respect to the response patterns in categorical response variables. Nonresponse can be a potential difficulty in applying LCM in practice, especially when missing values are nonignorable. Numerous techniques for handling nonresponse have been suggested for latent class models, but many of them require either MCAR or MAR assumptions which are not always acceptable. As one solution to the MNAR problem, we propose a selection model which uses an LCM to summarize nonresponse patterns into a smaller number of groups. Our proposed model is based on a latent ignorability assumption [15] which posits two categorical variables that summarize response patterns and nonresponse patterns on the questionnaire items, respectively.

We suggest two estimation methods for the proposed LCSM. Prospective analysts with large samples may consider using ML estimates via the EM algorithm, while analysts with specific prior information on their data (or small sample sizes) may choose to use Bayesian method. The EM algorithm provides stable estimates when appropriate initial values are given because the parameterization of the model provides an explicit form that can easily be maximized [17].

We articulate the EM algorithm details that are specifically applicable to our suggested model, and discuss the theoretical properties of the EM estimators to help prospective users draw appropriate inferences. Similarly, we propose an MCMC algorithm based on the Gibbs sampler with detailed steps and compare their performances with the proposed EM estimators as well as other conventional methods. Further, we have written a program to implement our proposed LCSM in the R language (version 3.6.1) that is available in *https://sites.google.com/site/leejwegg/*.

One drawback of the EM algorithm is its sensitivity to starting values, so using appropriate initial values was critical to obtain the global maximum [19]. To avoid the local maxima problem, users may try a large number of initial values and take the one with the highest likelihood as the final ML solution. The Bayesian estimation via MCMC algorithm, on the other hand, was not affected by the initial values due to the burn-in period, but it is exposed to a label-switching problem as suggested in [22]. For the various remedies for label-switching problems in latent class analysis, see [22, 29].

We analyze the 2018 GSS data using our proposed LCSM. Our proposed LCSM identified four response patterns of the abortion-attitude survey items and also discovered that the individuals' membership is related to their demographic factors such as age, gender, and race. In the proposed model, we assume that individual covariates only affect the latent response patterns, and such an assumption provides a simpler model with fewer parameters. Incorporating covariates effects on the latent missing propensity is straightforward. Our proposed model identified two sub-classes for each latent class where one represents the individuals who were likely to respond to all questionnaires (*completers*), and the other represents the individuals with nonresponse to at least one items (*incomplete*). Four (*complete*) subclasses showed very high probabilities of response to all questionnaire items thus are combined into a single subclass, but all (*incompleters*) subgroups had varying nonresponse patterns. Based on these findings, we conclude that our LCSM successfully identifies nonresponse patterns and their proportions.

In this paper, we focused on the simplest form of the latent class model that is confined to a data set collected within a single time point. Application of the latent class selection model can be extended to longitudinal data. For example, the latent class selection framework can be applied to identify the patterns of nonignorable dropouts. Also, we assumed that missingness is not affected by covariates (that is, $P(W \mid U)$ is independent of $\mathbf{X}_i$) to simplify the model. Including the effect of covariate on missingness via stratified multinomial logistic regression (i.e., $P(W = w \mid U = u, \mathbf{X}_i) = exp(\mathbf{X}_i\boldsymbol{\beta}_{w|u}) / \sum_{s=1}^{D} exp(\mathbf{X}_i\boldsymbol{\beta}_{s|u})$, for each u) can be a further research topic. Such extensions are expected to contribute to the specification of missingness under the MNAR mechanism.

APPENDIX I : SCORE FUNCTION OF LCSM

Eq. (18) shows the elements of first derivatives with respect to parameters $[\boldsymbol{\phi}, \boldsymbol{\rho}, \boldsymbol{\beta}, \boldsymbol{\delta}]$.

$$
\begin{aligned}
\frac{\partial logL}{\partial \rho_{mk|d}} &= \sum_{i=1}^{N} \frac{\theta_{i(d)} I(Y_{im} = k)}{\rho_{mk|d}}. \\
\frac{\partial logL}{\partial \phi_{ph|u,d}} &= \sum_{i=1}^{N} \frac{\theta_{i(u,d)} I(R_{ip} = h)}{\phi_{ph|u,d}}. \\
\frac{\partial logL}{\partial \delta_{u|d}} &= \sum_{i=1}^{N} \frac{\theta_{i(u,d)}}{\delta_{u|d}}. \\
\frac{\partial logL}{\partial \beta_{qd}} &= \sum_{i=1}^{N} X_{iq}(\theta_{i(d)} - \gamma_d(X_i)).
\end{aligned} \tag{18}
$$

Here, we have $p = 1 \ldots P$, $u = 1 \ldots S$, $d = 1, \ldots, D$, and $q = 1 \ldots K$. Also, some sets of parameters are constrained in a way that $\sum_{k=1}^{r_p} \rho_{mk|d} = 1$, $\sum_{h=0}^{1} \phi_{ph|u,d} = 1$, and $\sum_{u=1}^{S} \delta_{u|d} = 1$ for all subscripts. Suppose that $A_Q = [diag(1, Q-1), -\mathbf{1}_{Q-1}] \in \mathbb{R}^{Q-1 \times Q}$ denotes the constraint matrix for Q constrained parameters. Then, we can obtain the score function with respect to the free parameters as follows

$$S(\boldsymbol{\theta}) = A_Q f'(\boldsymbol{\theta})$$

where $f'(\boldsymbol{\theta})$ is the first derivative vector from Eq. (18).

APPENDIX II : HESSIAN MATRIX OF LCSM

The elements of the Hessian matrix with respect to each type of parameters $[\boldsymbol{\phi}, \boldsymbol{\rho}, \boldsymbol{\beta}, \boldsymbol{\delta}]$ are given next. Eq. (19) shows the second derivatives with respect to parameters $\boldsymbol{\phi}$ and the others.

(19)

$$\frac{\partial^2 logL}{\partial \rho_{mk|d} \partial \rho_{m'k'|d'}} = \sum_{i=1}^{N} \frac{\theta_{i(d)}(\zeta_{dd'}(1-\zeta_{pp'}) - \theta_{i(d')})\zeta_{Y_{im}k}\zeta_{Y_{im'}k'}}{\rho_{mk|d}\ \rho_{m'k'|d'}}$$

$$\frac{\partial^2 logL}{\partial \rho_{mk|d} \partial \phi_{ph|u,d'}} = \sum_{i=1}^{N} \frac{\theta_{i(u,d)}(\zeta_{dd'} - \theta_{i(u,d')})\zeta_{Y_{im}k}\zeta_{R_{ip}h}}{\rho_{mk|d}\ \phi_{ph|u,d'}}$$

$$\frac{\partial^2 logL}{\partial \rho_{mk|d} \partial \delta_{u|d'}} = \sum_{i=1}^{N} \frac{(\zeta_{dd'} - \theta_{i(d)})\theta_{i(u,d')}\zeta_{Y_{im}k}}{\rho_{mk|d}\ \delta_{u|d'}}$$

$$\frac{\partial^2 logL}{\partial \rho_{mk|d} \partial \beta_{qd'}} = \sum_{i=1}^{N} \frac{X_{iq}(\zeta_{dd'} - \theta_{i(d)})\theta_{i(d')}\zeta_{Y_{im}k}}{\rho_{mk|d}}$$

Here, we have $q = 1, \ldots, K,\ k = 0, 1,\ h = 1, \ldots, r_p,\ p = 1, \ldots, P,\ m = 1, \ldots, P,\ u = 1 \ldots S,\ d = 1 \ldots D$. Also, $\zeta_{dd'}$ is defined as 1 if $d = d'$ and 0 if $d \neq d'$

Eq. (20) shows the second derivatives with respect to parameters $\boldsymbol{\rho}$ and the others.

(20)

$$\frac{\partial^2 logL}{\partial \phi_{ph|u,d} \partial \phi_{p'h'|u',d'}} = \sum_{i=1}^{N} \frac{(\zeta_{uu'}\zeta_{dd'}(1-\zeta_{mm'}) - \theta_{i(u,d)})\theta_{i(u',d')}\zeta_{R_{ip}h}\zeta_{R_{ip'}h'}}{\phi_{ph|u,d}\ \phi_{p'h'|u',d'}}$$

$$\frac{\partial^2 logL}{\partial \phi_{ph|u,d} \partial \beta_{qd'}} = \sum_{i=1}^{N} \frac{X_{iq}\theta_{i(u,d)}(\zeta_{dd'} - \theta_{i(d')})\zeta_{R_{ip}h}}{\phi_{ph|u,d}}$$

$$\frac{\partial^2 logL}{\partial \phi_{ph|u,d} \partial \delta_{u'|d'}} = \sum_{i=1}^{N} \frac{\left\{\zeta_{dd'}\zeta_{uu'} - \theta_{i(u,d)}\theta_{i(u',d')}\right\}\zeta_{R_{ip}h}}{\phi_{ph|u,d}\ \delta_{u'|d'}}$$

Here, we have $q = 1, \ldots, K,\ k = 0, 1,\ h = 1, \ldots r_p,\ p = 1, \ldots P,\ m = 1, \ldots P,\ u = 1, \ldots S$ and $d = 1, \ldots D$.

Eq. (21) shows the second derivatives with respect to parameters $\boldsymbol{\delta}$ and the others.

$$\frac{\partial^2 logL}{\partial \delta_{u|d} \partial \delta_{u'|d'}} = -\sum_{i=1}^{N} \frac{\theta_{i(u,d)}\theta_{i(u',d')}}{\delta_{u|d}\ \delta_{u'|d'}}$$

(21)

$$\frac{\partial^2 logL}{\partial \delta_{u|d} \partial \beta_{qd'}} = \sum_{i=1}^{N} \frac{X_{iq}\theta_{i(u,d)}(\zeta_{dd'} - \theta_{i(d')})}{\delta_{u|d'}}$$

Here, we have $q = 1, \ldots, K,\ u = 1, \ldots, S$ and $d = 1, \ldots, D$.

Finally, Eq. (22) shows the second derivatives with respect to parameters $\boldsymbol{\beta}$.

(22)

$$\frac{\partial^2 logL}{\partial \beta_{qd} \partial \beta_{q'd'}} = \sum_{i=1}^{N} X_{iq}X_{iq'}\left\{\theta_{i(d)}(\zeta_{dd'} - \theta_{i(d')}) - \gamma_d(X_i)(\zeta_{dd'} - \gamma_{d'}(X_i))\right\}$$

Here, we have $q = 1 \ldots, K$ and $d = 1 \ldots D$.

Similar to the first derivatives, some sets of parameters are constrained in a way that $\sum_{k=1}^{r_p} \rho_{mk|d} = 1$, $\sum_{h=0}^{1} \phi_{ph|u,d} = 1$, and $\sum_{u=1}^{S} \delta_{u|d} = 1$ for all subscripts. We can obtain the Hessian matrix with respect to free parameters as follows

$$H(\boldsymbol{\theta}) = A_Q f''(\boldsymbol{\theta}) A_Q^T$$

where $f''(\boldsymbol{\theta})$ is the second derivative vector from Eq. (19), (20) and (21).

ACKNOWLEDGEMENTS

We thank the editor, associate editor, and reviewers for their critical assessment of our work.

Received 3 January 2019

REFERENCES

[1] H. Demirtas, J. L. Schafer, On the performance of random-coefficient pattern-mixture models for non-ignorable drop-out, Statistics in medicine 22 (16) (2003) 2553–2575.

[2] J. R. Cheema, Some general guidelines for choosing missing data handling methods in educational research, Journal of Modern Applied Statistical Methods 13 (2) (2014) 3.

[3] L. H. Rubin, K. Witkiewitz, J. S. Andre, S. Reilly, Methods for handling missing data in the behavioral neurosciences: Don't throw the baby rat out with the bath water, Journal of Undergraduate Neuroscience Education 5 (2) (2007) A71.

[4] N. R. Council, et al., The prevention and treatment of missing data in clinical trials (2010).

[5] J. J. Heckman, The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models, in: Annals of economic and social measurement, volume 5, number 4, NBER, 1976, pp. 475–492.

[6] R. J. Little, D. B. Rubin, Statistical analysis with missing data, Vol. 793, John Wiley & Sons, 2019.

[7] R. J. Little, Pattern-mixture models for multivariate incomplete data, Journal of the American Statistical Association 88 (421) (1993) 125–134.

[8] M. C. Wu, R. J. Carroll, Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process, Biometrics (1988) 175–188.

[9] P. Diggle, M. G. Kenward, Informative drop-out in longitudinal data analysis, Journal of the Royal Statistical Society: Series C (Applied Statistics) 43 (1) (1994) 49–73.

[10] H. Jung, J. L. Schafer, B. Seo, A latent class selection model for nonignorably missing data, Computational statistics & data analysis 55 (1) (2011) 802–812.

[11] M. G. Kenward, Selection models for repeated measurements with non-random dropout: an illustration of sensitivity, Statistics in medicine 17 (23) (1998) 2723–2732.

[12] A. L. McCutcheon, Latent class analysis, no. 64, Sage, 1987.

[13] H. Lin, C. E. McCulloch, R. A. Rosenheck, Latent pattern mixture models for informative intermittent missing data in longitudinal studies, Biometrics 60 (2) (2004) 295–305.

[14] J. Roy, Modeling longitudinal data with nonignorable dropouts using a latent dropout class model, Biometrics 59 (4) (2003) 829–836.

[15] O. Harel, J. L. Schafer, Partial and latent ignorability in missing-data problems, Biometrika 96 (1) (2009) 37–50.

[16] J. Kuha, M. Katsikatsou, I. Moustaki, Latent variable modelling with non-ignorable item non-response: multigroup response propensity models for cross-national analysis, Journal of the Royal Statistical Society: Series A (Statistics in Society) 181 (4) (2018) 1169–1192.
[17] K. Bandeen-Roche, D. L. Miglioretti, S. L. Zeger, P. J. Rathouz, Latent variable regression for multiple discrete outcomes, Journal of the American Statistical Association 92 (440) (1997) 1375–1386.
[18] G.-H. Huang, K. Bandeen-Roche, Building an identifiable latent class model with covariate effects on underlying and measured variables, Psychometrika 69 (1) (2004) 5–32.
[19] A. P. Dempster, N. M. Laird, D. B. Rubin, Maximum likelihood from incomplete data via the em algorithm, Journal of the Royal Statistical Society: Series B (Methodological) 39 (1) (1977) 1–22.
[20] N. M. Kiefer, Discrete parameter variation: Efficient estimation of a switching regression model, Econometrica: Journal of the Econometric Society (1978) 427–434.
[21] H. Chung, B. P. Flaherty, J. L. Schafer, Latent class logistic regression: application to marijuana use and attitudes among high school seniors, Journal of the Royal Statistical Society: Series A (Statistics in Society) 169 (4) (2006) 723–743.
[22] H. Chung, J. C. Anthony, A bayesian approach to a multiple-group latent class-profile analysis: the timing of drinking onset and subsequent drinking behaviors among us adolescents, Structural Equation Modeling: A Multidisciplinary Journal 20 (4) (2013) 658–680.
[23] L. M. Collins, J. L. Schafer, C.-M. Kam, A comparison of inclusive and restrictive strategies in modern missing data procedures., Psychological methods 6 (4) (2001) 330.
[24] J. W. Lee, H. Chung, A multivariate latent class profile analysis for longitudinal data with a latent group variable, Communications for Statistical Applications and Methods 27 (1) (2020) 15–35.
[25] H. R. F. Ebaugh, C. A. Haney, Shifts in abortion attitudes: 1972-1978, Journal of Marriage and the Family (1980) 491–499.
[26] E. J. Hall, M. M. Ferree, Race differences in abortion attitudes, Public Opinion Quarterly 50 (2) (1986) 193–207.
[27] C. Wilcox, Race differences in abortion attitudes: Some additional evidence, Public Opinion Quarterly 54 (2) (1990) 248–255.
[28] A. K. Formann, Linear logistic latent class analysis for polytomous data, Journal of the American Statistical Association 87 (418) (1992) 476–486.
[29] J. W. Lee, H. Chung, S. Jeon, Bayesian multivariate latent class profile analysis: Exploring the developmental progression of youth depression and substance use, Computational Statistics & Data Analysis 161 (2021) 107261.

Jung Wun Lee
215 Glenbrook Road, Storrs, CT, 06269
United States
E-mail address: jung.w.lee@uconn.edu

Ofer Harel
215 Glenbrook Road, Storrs, CT, 06269
United States
E-mail address: ofer.harel@uconn.edu