Running head: A BAYESIAN LATENT VARIABLE SELECTION MODEL

A Bayesian Latent Variable Selection Model for Nonignorable Missingness

Han Du, Craig Enders, Brian Keller, Thomas N. Bradbury, and Benjamin R. Karney

Department of Psychology, University of California, Los Angeles

Department of Educational Psychology, The University of Texas at Austin

Department of Psychology, University of California, Los Angeles

Department of Psychology, University of California, Los Angeles

Correspondence should be addressed to Han Du, Pritzker Hall, 502 Portola Plaza, Los Angeles, CA 90095.

Email: hdu@psych.ucla.edu.

This work was supported by Institute of Educational Sciences award R305D1900002.

Du, H., & Enders, C. K., Keller, B. T., Bradbury, T. N., & Karney, B. R. (2022). A Bayesian latent variable selection model for nonignorable missingness. Multivariate Behavioral Research, 57(2-3), 478-512.

Abstract

Missing data are exceedingly common across a variety of disciplines, such as educational, social, and behavioral science areas. Missing not at random (MNAR) mechanism where missingness is related to unobserved data is widespread in real data and has detrimental consequence. However, the existing MNAR-based methods have potential problems such as leaving the data incomplete and failing to accommodate incomplete covariates with interactions, non-linear terms, and random slopes. We propose a Bayesian latent variable imputation approach to impute missing data due to MNAR (and other missingness mechanisms) and estimate the model of substantive interest simultaneously. In addition, even when the incomplete covariates involves interactions, non-linear terms, and random slopes, the proposed method can handle missingness appropriately. Computer simulation results suggested that the proposed Bayesian latent variable selection model (BLVSM) was quite effective when the outcome and/or covariates were MNAR. Except when the sample size was small, estimates from the proposed BLVSM tracked closely with those from the complete data analysis. With a small sample size, when the outcome was less predictable from the covariates, the missingness proportions of the covariates and the outcome were larger, and the missingness selection processes of the covariates and the outcome were more MNAR and MAR, the performance of BLVSM was less satisfactory. When the sample size was large, BLVSM always performed well. In contrast, the method with an MAR assumption provided biased estimates and undercoverage confidence intervals when the missingness was MNAR. The robustness and the implementation of BLVSM in real data were also illustrated. The proposed method is available in the Blimp software application, and the paper includes a data analysis example illustrating its use.

A Bayesian Latent Variable Selection Model for Nonignorable Missingness

Missing data are exceedingly common across a variety of disciplines such as the educational, social, and behavioral sciences. Participants drop out of studies or omit responses for a variety of reasons, some of which are benign, but others of which can have serious consequences on the validity of a statistical analysis if the missing values aren't dealt with properly. Mainstream missing data handling methods typically assume a missing at random (MAR) mechanism, whereby the probability of missingness is only related to observed scores (Little & Rubin, 2014). For example, students could opt out of achievement testing for reasons related to background variables such as socioeconomic status, language proficiency but not to achievement itself. The MAR assumption is reasonable in many cases, but there are also many situations where missingness is related to unobserved scores themselves (Little & Rubin, 2014). This type of missingness process is called a missing not at random mechanism (MNAR; also known as nonignorable missingness). For example, in education, students with low achievement could have missing values on an achievement test because they fail to finish the exam. Hence, the missingness of the achievement scores is due to the unobserved ability. In medical trial settings, values of physical measures could be missing because patients die during the treatment period. Therefore, the missingness of physical measure scores is due to the unobserved physical status, even after conditioning on the observed data. Additionally, in substance use cessation studies, participants may skip a blood or urine test because they are using substances and will have positive test results. In this case, the missingness of test results is directly related to the unobserved test results.

If the true underlying missingness mechanism is MNAR but an MAR-based analysis procedure is used, previous research has shown that parameter estimates will generally be biased (Enders, 2011; Fitzmaurice et al., 2012; Graham, 2009; Yang & Maxwell, 2014). The fundamental problem is that it is difficult to fully rule out the possibility of MNAR mechanism because the observed data carry no information about the unobserved scores and their associations with other variables. This makes correcting

for MNAR inherently complex because missingness depends on the unobserved information. In practice, it is necessary to simultaneously estimate the analysis model of substantive interest and an additional model for the nonresponse process (e.g., a regression model where the outcome or covariate predicts its own binary missing data indicator). In other words, an MNAR mechanism requires that we model the joint distribution of the data and missingness, $p(y, r)$. In our generic notation, $p(y)$ represents the distribution induced by the *substantive model* (e.g., a linear regression model) and $p(r)$ denotes the corresponding distribution of *missingness model* where r is the missing data indicator. In principle, MNAR processes can apply to the outcome or predictors in a substantive model. Existing literature focuses on the nonignorable missingness on the outcome, except that Ibrahim et al. (1999) and Ibrahim et al. (2005) briefly showed how to handle nonignorable covariates. This paper is the first one which presents all combinations of missingness mechanisms, whereas the previous literature focuses on missing outcome or missing covariates separately. Additionally, there are three distinctions between Ibrahim's work and our work, which will be elaborated later. As a brief preview, our proposed Bayesian procedure uses probit regression with latent variables to model missingness.

There are two broad MNAR modeling frameworks: the selection model and pattern mixture model. Heckman (1976; 1979) originally proposed the selection model for an MNAR process on the outcome, and Glynn et al. (1986), Little (1993; 1994), and Rubin (1987) proposed the general form of the pattern-mixture model. The two frameworks both integrate a missingness model that captures the propensity for missing data in the analysis, but they factorize the joint distribution and operationalize the missingness model differently. The selection model framework partitions the joint distribution of the data and missingness as $p(y, r) = p(y)\, p(r|y)$. The second term, $p(r|y)$, says that missingness is modeled as a function of the incomplete variable y (Heckman, 1976, 1979). As noted previously, this representation often entails the simultaneous estimation of two models: a regression model for the outcome of substantive interest, and a second model with y's missing data indicator as a function of y and possibly other variables.

In contrast, the pattern-mixture model framework partitions the joint distribution as $p(y, r) = p(r)\, p(y|r)$. The second term, $p(y|r)$, describes how the model of substantive interest depends on the missing data pattern (e.g., Little, 1993). This representation reverses the role of r, such that the substantive analysis model parameters vary across missing data patterns. In this framework, the model of substantive interest can be estimated separately for each missing data pattern (usually with a set of identification constraints), or the missing data patterns can appear as dummy-coded covariates in the analysis (Hedeker & Gibbons, 1997). Ibrahim et al. (1999), Huang et al. (2005), and Ibrahim et al. (2005) provided and summarized methods to handle MNAR in generalized linear models. The former two focused on the selection model, and the latter discussed the pattern mixture model. Galimard et al. (2016) and Galimard et al. (2018) extended selection models to multiple incomplete covariates in the chained equations framework. In addition, a good deal of methodological research has developed variants of these approaches for longitudinal data. For example, Diggle & Kenward (1994) outlined a selection model for longitudinal data analyses. P. S. Albert & Follmann (2000), Follmann & Wu (1995), Wu & Bailey (1989), and Wu & Carroll (1988) proposed another type of longitudinal selection model called a random coefficient selection model (also referred to as the shared parameter approach) whereby random effects predict missingness. Extending the Diggle & Kenward selection model, Daniels & Hogan (2008) proposed a Bayesian selection model when longitudinal outcomes are missing. Within the pattern-mixture model framework, Roy (2003) introduced a pattern-mixture method treating class membership as a latent variable. Other applications have combined features of pattern-mixture model and selection model or have otherwise developed variants of the two frameworks (e.g., Beunckens et al., 2008; Dantan et al., 2008; Demirtas & Schafer, 2003; Foster et al., 2004; Galimard et al., 2016; Gottfredson et al., 2014; Hafez et al., 2015; Roy & Daniels, 2008; Yuan & Little, 2009; Mason et al., 2012; Muthén et al., 2011). It is important to emphasize that the selection and pattern mixture models are not exchangeable representations of the joint distribution. For example, we would not expect pattern mixture models to accurately capture a process aligned with

$p(y, r) = p(y)\, p(r|y)$, nor would we expect our method to yield unbiased estimates if the true process is $p(y, r) = p(r)\, p(y|r)$. This is an inherent feature of MNAR modeling and not Bayesian estimation, per se.

The major limitation of existing MNAR methods is that they focus on incomplete outcomes and don't necessarily provide a mechanism for handling MNAR explanatory variables except the expectation–maximization (EM) algorithm proposed by Ibrahim et al. (1999) and Ibrahim et al. (2005). Recent studies have described fully Bayesian estimation and imputation approaches that allow for MAR covariates with interactions, non-linear terms, and random slopes (e.g., Bartlett et al. 2015; Enders et al. Advance online publication; Erler et al. 2019, 2016; Goldstein et al. 2014; Kim et al. 2015, 2018; Lüdtke et al. 2020; Zhang & Wang 2017). We refer to these methods generically as *model-based estimation and imputation* because they essentially tailor missing values to the substantive model of interest. These approaches yield Bayesian estimates of the model parameters, and they also generate imputations that can be analyzed in the frequentist framework with Rubin's pooling rules (Rubin, 1987). As mentioned previously, our approach readily accommodates an MNAR process for any variable in the model including covariates, and thus it is a generalization of existing model-based imputation.

The purpose of this study is to outline a fully Bayesian latent variable selection model (BLVSM) to impute missing data and estimate parameters of interest where either covariates or outcomes follow an MNAR (or MAR) process. The method falls in the class of selection models outlined by Heckman and others (Diggle & Kenward, 1994; Heckman, 1976, 1979; Huang et al., 2005; Ibrahim et al., 1999), and we apply a Bayesian estimation procedure that simultaneously estimates the substantive regression model and a probit regression model that invokes latent missingness variables. Besides direct Bayesian inference, multiple imputations are a natural by-product of the Markov Chain Monte-Carlo (MCMC) estimation algorithm. These imputations can be analyzed in the frequentist framework in lieu of direct Bayesian inference to answer various of new research questions, without requiring any special software. Our approach can accommodate general missing data patterns and the following scenarios: the outcome is

MAR or MNAR with a) complete covariates, b) incomplete covariates with an MAR mechanism, and c) incomplete covariates with MNAR mechanism. It also can be applied when the outcome is complete but covariates are MAR or MNAR. Importantly, we extend the work in Enders et al. (Advance online publication), where the substantive analysis model supports incomplete MAR non-linear functions such as interactive and polynomial effects. This model-based estimation and imputation procedure extends to accommodate incomplete MNAR covariates with a variety of metrics (continuous, binary, ordinal, or nominal). The proposed procedure is ready in a forthcoming release of the software Blimp 3 (Keller et al., 2019). We are unaware of existing approaches and software that can handle these combinations of features, although the R package 'mdmb' can estimate some selection models.

The outline of this paper is as follows: in "A Typical Selection Model" section, an overview of selection model is given. In "Bayesian Estimation of a Selection Model with MAR Covariates" and "Bayesian Estimation of a Selection Model with MNAR Covariates" sections, we present the proposed fully Bayesian latent variable selection model (BLVSM) when covariates are MAR and MNAR, respectively. In "Simulation Study 1: MAR Covariates", "Simulation Study 2: MNAR Covariates", and "Simulation Study 3: Misspecification" sections, the performance of BLVSM when covariates are MAR and MNAR and when the selection model is misspecified is thoroughly examined via simulations, respectively. In "A Real Data Example" section, a real data example is provided to illustrate BLVSM. We end the paper with some concluding remarks in "Conclusion" section.

A Typical Selection Model

In this paper, we consider the case where the missingness is a function of the unseen scores and possibly other variables. To illustrate the missing data handling procedure for a MNAR mechanism, we start by focusing on an incomplete outcome. We consider a simple regression model where y has missing values and missingness is a function of the y scores themselves. As illustrated earlier, a selection model

consists of two components: the substantive model $p(y)$ and the missingness model $p(r|y)$.

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad \varepsilon_i \sim N\left(0, \sigma_\varepsilon^2\right) \tag{1}$$

$$r_{yi}^* = \gamma_0 + \gamma_1 y_i + \zeta_i \qquad \zeta_i \sim N(0, 1)$$

The first part of Equation (1) presents the substantive model $p(y)$. We introduce a binary missing data indicator r_i, where $r_i = 0$ if y_i is observed and $r_i = 1$ if y_i is missing. The second part of Equation (1) presents the missingness model $p(r|y)$, which is a probit regression model defining missingness as a normally distributed latent variable (Johnson & Albert, 2006). r_{yi}^* is a continuous latent missingness variable for individual i that represents an individual's latent propensity or proclivity for missing data. For example, in an education context, y_i could be an achievement test score that is potentially missing for reasons related to achievement ability itself (i.e., a student may fail to finish the exam and thus it leads a missing value), and r_{yi}^* represents how likely a student fail to finish the exam. The fixed part of the missingness model, $\gamma_0 + \gamma_1 y_i$, defines the conditional mean (expected value) of the latent variable. In other words, the fixed part defines the systematic influence of missingness due to the unobserved outcome scores. The residual ζ_i is standard normal with variance fixed at one for identification. Accordingly, the probit regression can be written as

$$Pr(r_i = 1|y_i) = 1 - \Phi(\gamma_0 + \gamma_1 y_i), \tag{2}$$

where $\Phi()$ is the cumulative distribution function of the standard normal distribution. $\gamma_0 + \gamma_1 y_i$ is the predicted z-score of missingness propensity and $\Phi(\,)$ returns the proportion of the area below that z score in a standard normal curve. The probit regression model additionally incorporates a threshold parameter κ that divides the standard normal latent distribution into two segments, such that $r_i = 0$ if $r_{yi}^* < \kappa$ and $r_i = 1$ if $r_{yi}^* \geq \kappa$. That is, the latent missingness scores increase to a cut-point, above which the score

becomes missing. Note that κ is typically fixed at zero to avoid redundancy with the regression intercept, but the model can also be parameterized by eliminating the intercept and estimating the threshold. Because estimating the threshold parameter is known to exhibit slow convergence behavior (Cowles, 1996), we adopt the former approach and fix the threshold at zero.

Bayesian Estimation of BLVSM with MAR Covariates

In this section, we describe the MCMC estimation steps for the case where the outcome in the substantive model is MNAR and the covariates are MAR. The Bayesian framework views the substantive regression model parameters, missingness model parameters, missing outcome scores, covariate model parameters, and missing covariate scores as variables to be estimated. The Gibbs sampler breaks this complex multivariate problem involving parameters and missing values into a series of simpler univariate problems, each of which draws one of the unknown quantities at random from a probability distribution that conditions on the current values of all other unknowns, which will be elaborated later (Gelfand & Smith, 1990). After providing the posterior distribution of each variable, we illustrate how to estimate each variable by Gibbs sampling procedure.

To illustrate our proposed Bayesian latent variable selection model (BLVSM), we consider a single-level substantive model with multiple covariates,

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \varepsilon, \tag{3}$$

where $\boldsymbol{y}$ is a $N \times 1$ vector of the outcome measures for N individuals, $\boldsymbol{X}$ is a $N \times (K+1)$ matrix for the K covariates and one intercept, $\boldsymbol{\beta}$ is a $(K+1) \times 1$ vector for the $K+1$ regression coefficients, ε is a $N \times 1$ vector for independently distributed errors, and $\varepsilon \sim N\left(\mathbf{0}, \sigma_{\varepsilon}^{2}\boldsymbol{I}\right)$. As noted previously, our model specification readily accommodates incomplete interactive or curvilinear effects (Enders et al., Advance online publication), and it thus extends recent research (Bartlett et al., 2015; Du et al., Manuscript

submitted for publication; Erler et al., 2016; Goldstein et al., 2014; Grund et al., 2018; Ibrahim et al., 2002; Kim et al., 2015, 2018; Zhang & Wang, 2017) by accommodating an MNAR process for the outcome and/or an MNAR process for the covariates (presented in the next section). This combination of features is a new contribution to the literature, although researchers have worked on MAR covariates or MNAR outcome separately. To this end, consider the following moderated regression model, examples of which are exceedingly common in the literature (Aiken et al., 1991),

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + \varepsilon_i, \\ \varepsilon_i &\sim N\left(0, \sigma_\varepsilon^2\right). \end{aligned} \tag{4}$$

In this case

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_{1,1} & x_{1,2} & x_{1,1}x_{1,2} \\ 1 & x_{2,1} & x_{2,2} & x_{2,1}x_{2,2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{N,1} & x_{N,2} & x_{N,1}x_{N,2} \end{pmatrix}, \ \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}.$$

The missingness of y is a function of the y scores themselves and x_1 is incomplete due to an MAR process, and hence $x_1 x_2$ is incomplete. Indeed, regardless of whether the covariates are complete or incomplete, the posterior distributions of substantive model parameters, missingness model, and missing outcome are not affected.

As mentioned above, when a missing outcome, y_i, is related to the unobserved scores (e.g., y_i itself) for individual i, we introduce a binary missing data indicator, r_i, for which 1 indicates a missing outcome and 0 indicates an observed outcome. We generalize Equation (1) that the missingness not only depends on the unobserved y_i but also other variables to provide a more general form. Accordingly, an underlying

continuous latent missingness variable r^*_{yi} could be directly modeled by a regression model

$$\boldsymbol{r^*_y} = \boldsymbol{Z\gamma} + \boldsymbol{\zeta}, \tag{5}$$

where $\boldsymbol{r^*_y}$ is a $N \times 1$ vector of latent missingness propensities for N individuals, $\boldsymbol{Z} = (\boldsymbol{1}, \boldsymbol{y}, \boldsymbol{M})$ is a $N \times (2 + P)$ matrix, $\boldsymbol{M}$ is a $N \times P$ matrix for causes of missingness other than y itself, such as auxiliary variables, $\boldsymbol{\gamma}$ is a $(2 + P) \times 1$ vector, and $\boldsymbol{\zeta}$ is a $N \times 1$ vector of error term with $\boldsymbol{\zeta} \sim N(\boldsymbol{0}, \boldsymbol{I})$. Past literature suggests that if $\boldsymbol{M}$ incorporate the predictors in the substantive model $\boldsymbol{X}$ or other predictors highly correlated with $\boldsymbol{y}$, collinearity problems may occur and be detrimental to estimation (for details refer to Puhani, 2000; Stolzenberg & Relles, 1990, 1997). As discussed later, this does not appear to be the case with BLVSM, and we will recommend including variables from the substantive model in the missingness model. Latent variable scores r^*_{yi} follow a truncated normal distribution, such that r^*_{yi} is above the threshold ($\kappa = 0$) if y_i is missing (i.e., $r_i = 1$) and below the threshold if y_i is complete (i.e., $r_i = 0$).

Posterior Distributions of Substantive Model Parameters

We estimate the aforementioned selection model using an iterative MCMC algorithm, Gibbs sampling, that draws each unknown from a probability distribution that conditions on all other unknowns. The remainder of this section gives the full conditional distributions for the estimation steps. To begin, augmenting the likelihood with the latent missingness variable $\boldsymbol{r^*_y}$ gives

$$\begin{aligned} p\left(\boldsymbol{y}, \boldsymbol{r^*_y}|\boldsymbol{\gamma}, \boldsymbol{\beta}, \sigma^2_\varepsilon\right) &= p\left(\boldsymbol{y}|\boldsymbol{\beta}, \sigma^2_\varepsilon\right) p\left(\boldsymbol{r^*_y}|\boldsymbol{\gamma}, \boldsymbol{y}, \boldsymbol{M}\right) \\ &= \left(2\pi\sigma^2_\epsilon\right)^{-\frac{N}{2}} exp\left(-\frac{(\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta})}{2\sigma^2_\varepsilon}\right) \times (2\pi)^{-\frac{N}{2}} exp\left(-\frac{\left(\boldsymbol{r^*_y} - \boldsymbol{Z\gamma}\right)'\left(\boldsymbol{r^*_y} - \boldsymbol{Z\gamma}\right)}{2}\right), \end{aligned} \tag{6}$$

where $\boldsymbol{Z} = (\boldsymbol{1}, \boldsymbol{y}, \boldsymbol{M})$. We employ independent priors that $p(\beta) \propto 1$ for all coefficients in $\boldsymbol{\beta}$, $p\left(\sigma^2_\epsilon\right) = IG(a, a)$, and $p(\gamma) = N(0, b)$ for all coefficients in $\boldsymbol{\gamma}$. Note that $p(\gamma)$ needs to be weakly

informative, as some prior information is often needed to facilitate convergence, particularly in small samples. We will elaborate this point later in Simulation Study 1.

Based on the priors and the likelihood (Equation 6), the joint posterior distribution of $\boldsymbol{\gamma}$, $\boldsymbol{\beta}$, and σ_{ε}^2 is constructed by Bayes' theorem,

$$\begin{aligned} p\left(\boldsymbol{\gamma}, \boldsymbol{\beta}, \sigma_{\epsilon}^2 | \boldsymbol{y}, \boldsymbol{r}_{\boldsymbol{y}}^*\right) &\propto p\left(\boldsymbol{y}, \boldsymbol{r}_{\boldsymbol{y}}^* | \boldsymbol{\gamma}, \boldsymbol{\beta}, \sigma_{\varepsilon}^2\right) p\left(\boldsymbol{\gamma}, \boldsymbol{\beta}, \sigma_{\varepsilon}^2\right) \\ &\propto \left(\sigma_{\epsilon}^2\right)^{-\frac{N}{2}-a-1} exp\left(-\frac{2a + (\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})}{2\sigma_{\varepsilon}^2}\right) \times \\ &\quad exp\left(-\frac{\left(\boldsymbol{r}_{\boldsymbol{y}}^* - \boldsymbol{Z}\boldsymbol{\gamma}\right)'\left(\boldsymbol{r}_{\boldsymbol{y}}^* - \boldsymbol{Z}\boldsymbol{\gamma}\right)}{2}\right) \times exp\left(-\frac{\boldsymbol{\gamma}'\boldsymbol{\gamma}}{2b}\right) \end{aligned} \tag{7}$$

Based on the joint posterior distribution of $\boldsymbol{\gamma}$, $\boldsymbol{\beta}$, and σ_{ε}^2, we can derive the conditional posterior distributions one by one. Specifically, the conditional posterior distribution of $\boldsymbol{\beta}$ is a multivariate normal distribution with $\cdot$ indicating the variables and parameters conditional on

$$p\left(\boldsymbol{\beta}|\cdot\right) = MN\left(\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{y}, \sigma_{\varepsilon}^2\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right). \tag{8}$$

The conditional posterior distribution of σ_{ϵ}^2 is an Inverse-Gamma distribution,

$$p\left(\sigma_{\epsilon}^2|\cdot\right) = IG\left(\frac{N}{2} + a, \frac{(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})}{2} + a\right). \tag{9}$$

In words, Equation (8) says that the substantive model's regression coefficients are drawn from a multivariate normal distribution. The center and spread of this distribution align with ordinary least squares estimates of the coefficients and their parameter covariance matrix, respectively, because of the specified prior. Equation (9) says that the substantive model's residual variance is drawn at random from a right-skewed inverse gamma distribution. The center and spread of this distribution is determined by the degrees of freedom, residual sum of squares, and prior information. Note that the conditional posterior

distributions of $\boldsymbol{\beta}$ and σ_ϵ^2 are exactly the same as those from any linear regression problem using the same priors regardless of whether the outcome and covariates are incomplete.

Posterior Distributions of Missingness Model Parameters and Missing Outcome

Given the latent missingness propensity r^*, the coefficients of the missingness model have a posterior with a similar form as $\boldsymbol{\beta}$. That is, the MCMC algorithm draws a vector of regression coefficients from a multivariate normal distribution. The residual variance is not an estimated parameter here, as it is fixed at 1.

$$p\left(\boldsymbol{\gamma}|\cdot\right) = MN\left(\hat{\boldsymbol{\gamma}} = \Sigma_1^{-1}\boldsymbol{Z'r_y^*}, \Sigma_1 = \left(\frac{1}{b} \times I + \boldsymbol{Z'Z}\right)^{-1}\right). \tag{10}$$

In words, Equation (10) says that the selection model's regression coefficients are drawn from a multivariate normal distribution. The center and variance are determined by the latent data and current imputed data. All that is left is to define the distributions of the latent variable scores r_y^* and the missing values of y. Latent variable scores can be modeled by a truncated normal distribution,

$$p\left(r_{yi}^*|\cdot\right) = \begin{cases} N\left(\boldsymbol{Z_i\gamma}, 1\right) \mathrm{I}\left(\kappa = 0, \infty\right) & r_i = 1 \\ N\left(\boldsymbol{Z_i\gamma}, 1\right) \mathrm{I}\left(-\infty, \kappa = 0\right) & r_i = 0 \end{cases}, \tag{11}$$

where I denotes the indicator function, $\boldsymbol{Z_i}$ denotes the ith row of $\boldsymbol{Z}$ for individual i, and the threshold κ is fixed at 0. In words, this equation says that latent missingness scores are drawn from one of two normal distributions, both of which are centered at the predicted z-score from the regression equation (the mean of the normal distribution) and have a fixed variance equal to 1. Specifically, if y_i is observed, a latent score should be drawn from the region of the normal curve below 0 (the fixed threshold parameter), as this area corresponds to the region occupied by indicator scores of $r = 0$. Otherwise, if y_i is missing, a latent score

should be drawn from the region of the normal curve above the threshold, as this area corresponds to the region occupied by indicator scores of $r = 1$.

The posterior predictive distribution of the missing outcome values has a complex mean and variance structure that depends on the parameters of both the substantive model and missingness model. Conceptually, the mean of the normal distribution is a predicted value, but that prediction accounts for y's role as an outcome in the substantive model and a predictor in the selection model. The variance of the imputations similarly depends on terms from both models. Specifically, the posterior predictive distribution is proportional to the augmented likelihood in Equation (6) [1]

$$\begin{aligned} p\left(\boldsymbol{y}|\boldsymbol{r}^*_{\boldsymbol{y}}, \boldsymbol{\gamma}, \boldsymbol{\beta}, \sigma^2_\varepsilon\right) &\propto p\left(\boldsymbol{y}, \boldsymbol{r}^*_{\boldsymbol{y}}|\boldsymbol{\gamma}, \boldsymbol{\beta}, \sigma^2_\varepsilon\right) \\ &= MN\left(\frac{\boldsymbol{X\beta} + \gamma_y\sigma^2_\varepsilon\left(\boldsymbol{r}^*_{\boldsymbol{y}} - \boldsymbol{Z}_{-y}\boldsymbol{\gamma}_{-y}\right)}{1 + \gamma^2_y\sigma^2_\varepsilon}, \frac{\sigma^2_\varepsilon}{1 + \gamma^2_y\sigma^2_\varepsilon}\boldsymbol{I}\right), \end{aligned} \tag{12}$$

where γ_y is the regression coefficient for y in the missingness model, $\boldsymbol{\gamma_{-y}}$ are the regression coefficients except for y in the missingness model, and $\boldsymbol{Z_{-y}}$ are the predictors in the missingness model except y. Because Equation (12) is tedious to derive, alternatively, we can use the Metropolis-Hastings algorithm to empirically construct the posterior distribution and estimate the missing outcome scores from this distribution (Gilks et al., 1996; Hastings, 1970). The Metropolis-Hastings algorithm can used to draw the posterior samples of other parameters (i.e., $\boldsymbol{\beta}$, σ^2_ε, and $\boldsymbol{r}^*_{\boldsymbol{y}}$). Please see the supplemental materials for more information for the Metropolis-Hastings algorithm.

Posterior Distributions of Missing Covariates and Covariate Model Parameters

We assume that some of the covariates in the substantive model are partially observed and that missingness for the covariates depends on the fully observed covariates, the outcome, and/or other auxiliary variables. Suppose there are Q partially observed predictors (i.e., $x_1, ..., x_Q$) and $K - Q$ fully

observed predictors (i.e., $x_{Q+1}, ..., x_K$). We factorize the joint distribution of all incomplete covariates as

$$p\left(y, x_1, ..., x_Q | x_{Q+1}, ..., x_K\right) = p\left(y | x_1, ..., x_K\right) p\left(x_1, ..., x_Q | x_{Q+1}, ..., x_K\right),$$

where $p\left(y, x_1, ..., x_Q | x_{Q+1}, ..., x_K\right)$ is the joint distribution for all incomplete covariates and the outcome, $p\left(y | x_1, ..., x_K\right)$ is the distribution of y induced by the substantive model (i.e., a normal distribution, conditional on the covariates and possibly curvilinear or interactive terms), and $p\left(x_1, ..., x_Q | x_{Q+1}, ..., x_K\right)$ is the joint distribution of the incomplete covariates and represents the *covariate model* (e.g., a normal distribution for continuous and latent categorical covariates). We assume that the conditional distribution of the incomplete covariate variables, $p\left(x_1, ..., x_Q | x_{Q+1}, ..., x_K\right)$, is a multivariate normal distribution, such that the incomplete covariates are linearly related. Based on this assumption, we can specify the full conditional distribution for each incomplete covariate given all other incomplete and complete covariates as a univariate normal distribution. This is the so-called *"separate" specification* or *fully conditional specification* for covariates (Du et al., Manuscript submitted for publication; Enders, in press; Enders et al., Advance online publication). Alternatively, a *"sequential" specification* of the joint distribution (Erler et al., 2016, 2019; Ibrahim et al., 2002; Lüdtke et al., 2020) can accommodate non-linear relations among incomplete covariates, and this approach is equivalent to the separate specification when assuming multivariate normality (for more details, please refer to Du et al., Manuscript submitted for publication). Due to the scope and word limitation of this paper, we illustrate details of the sequential specification when covariates are MAR and outcome is MNAR in the supplemental materials. The focus of the main text is the separate or fully conditional specification because the separate specification is easier to implement and calculate especially for applied researchers. It is generally harder to implement the sequential specification because the researcher needs to work out how to factorize the joint distribution to achieve the desired model. Under a separate specification, the researcher just needs to specify the needed univariate covariate

model and nothing else. Our software Blimp can accommodate either specification because the sequential specification is an important alternate and is the only option when researchers would like to model nonlinear relations between covariates.

In the moderated regression example (Equation 4), to impute x_1 (or any of the covariates),we must derive its conditional distribution given all of the other variables including the outcome. Generally, we denote that x_q ($q = 1, ..., Q$; e.g., x_1 in Equation (4)) is the target of imputation at a particular set, and $\boldsymbol{x_{-q}}$ is set of all remaining covariates including the complete covariates except x_q, that is $\boldsymbol{x_{-q}} = \left\{x_1, ..., x_{(q-1)}, x_{(q+1)}, ..., x_Q, x_{Q+1}, ..., x_K\right\}$ (e.g., $x_{-1} = \{x_2\}$ in Equation (4)). $p\left(x_q|\boldsymbol{x_{-q}}\right)$ is a linear regression of x_q on all other covariates which is the covariate model, and $p\left(y|x_q, \boldsymbol{x_{-q}}\right)$ is the distribution of y induced by the substantive model (e.g., Equation (3)). We refer to this as the "separate" specification because each incomplete predictor requires a unique regression. Because x_q appears in both terms on the right side of the substantive model and on the left side of the covariate model, its posterior distribution has a complex form that depends on the product of two normal distributions. The resulting distribution of x_q given all other variables is (Du et al., Manuscript submitted for publication; Enders et al., Advance online publication; Erler et al., 2016; Kim et al., 2015)

$$p\left(x_q|y, \boldsymbol{x_{-q}}\right) \propto p\left(y|x_q, \boldsymbol{x_{-q}}\right) p\left(x_q|\boldsymbol{x_{-q}}\right). \tag{13}$$

In words, Equation (13) says that the distribution of x_q given all other variables depends on the distribution of y induced by the substantive model (x_q is a covariate in that model) and a normal distribution induced by the regression of x_q on all other predictors (i.e., the covariate model). Deriving the distribution of missing values involves multiplying all distributions that feature x_q then performing algebra that combines the component distributions into a single function of that covariate. We give the distribution below in

Equation (16). Particularly, the covariate model of x_q is $p\left(\boldsymbol{x_q}|\boldsymbol{X_{-q}}, \boldsymbol{\psi_q}, \sigma^2_{e,q}\right)$,

$$\boldsymbol{x_q} = \boldsymbol{X_{-q}}\boldsymbol{\psi_q} + \boldsymbol{e_q}, \tag{14}$$

where $\boldsymbol{x_q}$ is a $N \times 1$ vector of the target of imputation covariate for N individuals, $\boldsymbol{X_{-q}} = (\mathbf{1}, \boldsymbol{X_{Inc,-q}}, \boldsymbol{X_{obs}})$ is a $N \times K$ matrix (e.g., $\boldsymbol{X_{-1}} = (\mathbf{1}, \boldsymbol{X_2})$ in Equation (4)), $\boldsymbol{X_{Inc,-q}}$ denotes all the incomplete covariates except x_q for N individuals, $\boldsymbol{X_{obs}} = (\boldsymbol{x_{Q+1}}, ..., \boldsymbol{x_K})$ denotes all the observed covariate for N individuals, $\boldsymbol{\psi_q}$ is a $K \times 1$ vector for for K regression coefficients, $\boldsymbol{e_q}$ is a $n \times 1$ vector, and $\boldsymbol{e_q} \sim N\left(\mathbf{0}, \sigma^2_{e,q}\boldsymbol{I}\right)$. When assuming that the missing predictor is conditional on auxiliary variables, the auxiliary variables enter Equation (14) as predictors. When using the separate specification, it implies that the incomplete covariates follow a multivariate normal distribution and thus restrict incomplete covariates to be linearly related. Therefore, $\boldsymbol{X_{-q}}$ cannot contain curvilinear or interaction terms involving any incomplete covariates, although the incomplete curvilinear and interaction terms can appear in the substantive model. If one prefers to use the sequential specification for the covariate model, the details are provided in the supplemental materials (e.g., Erler et al., 2016; Lüdtke et al., 2020).

Note that we assume the missingness of the outcome does not depend on the unobserved values of $\boldsymbol{x_q}$ in the above model. If the missing outcome is not only related to the unobserved outcome itself but also conditional on the incomplete covariates in the substantive model (e.g., $\boldsymbol{M} = \boldsymbol{X}$). Then the posterior distribution of x_q should consider its influence on the underlying continuous latent missingness variable of the outcome r^*_y by

$$p\left(\boldsymbol{x_q}|\cdot\right) \propto p\left(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\beta}, \sigma^2_\varepsilon\right) p\left(\boldsymbol{x_q}|\boldsymbol{X_{-q}}, \boldsymbol{\psi_q}, \sigma^2_{e,q}\right) p\left(\boldsymbol{r^*_y}|\gamma, \boldsymbol{y}, \boldsymbol{x_q}, \boldsymbol{X_{-q}}\right). \tag{15}$$

But this missingness assumption may cause a collinearity problem.

Back to the moderated regression example in Equation (4), there is no need to specify a model for

x_1x_2, as the lower-order terms are sampled from a distribution that accounts for their role in the product. Kim et al. (2015) show that estimating the lower-order scores in this fashion is equivalent to sampling x_1 and the x_1x_2 product as a pair. We estimate x_1 by Equation (13) and compute x_1x_2 based on the imputed x_1 and x_2. The covariate model for x_1, $p\left(x_1|x_2\right)$, is defined as a linear regression $x_{1i} = \psi_0 + \psi_1 x_{2i} + e_i$ with $e_i \sim N\left(0, \sigma_e^2\right)$. The posterior distribution of x_1 based on Equation (13) is

$$
\begin{aligned}
p\left(x_{1i(miss)}|y_i, x_{2i}\right) &\propto p\left(y_i|x_{1i}, x_{2i}\right) p\left(x_{1i}|x_{2i}\right) \\
&= N\left(\frac{\sigma_e^2\left(\beta_1 + \beta_3 x_{2i}\right)\left(y_i - \beta_0 - \beta_2 x_{2i}\right) + \sigma_\varepsilon^2\left(\psi_0 + \psi_1 x_{2i}\right)}{\sigma_e^2\left(\beta_1 + \beta_3 x_{2i}\right)^2 + \sigma_\varepsilon^2}, \frac{\sigma_\varepsilon^2\sigma_e^2}{\sigma_e^2\left(\beta_1 + \beta_3 x_{2i}\right)^2 + \sigma_\varepsilon^2}\right).
\end{aligned}
\tag{16}
$$

The distribution of x_q imputations is a normal distribution, albeit a complicated one with a mean and variance that depend on two sets of model parameters. The mean is a function of the substantive model's parameters as well as the covariate model's parameters. The variance similarly depends on two models (note that the variance is heteroscedastic and depends on a participant's moderator score). As an aside, a result similar to Equation (16) cannot be derived for incomplete curvilinear effects (e.g., $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i}^2 + \varepsilon_i$ with incomplete x_1) (Lüdtke et al., 2020). However, the absence of an analytical form for this posterior distribution is not problematic in practice, as we can use the Metropolis-Hastings algorithm to estimate the missing covariates. This strategy provides a more general solution across a variety of analytic scenarios, including the ones we examine here. Note that the procedure is not limited to a single incomplete covariate.

For the posterior distributions of coefficients in the covariate model, $\boldsymbol{\psi_q}$ and $\sigma_{e,q}^2$, we employ a Jeffreys prior where $p\left(\psi\right) \propto 1$ for all coefficients in $\boldsymbol{\psi_q}$ and $p\left(\sigma_{e,q}^2\right) \propto \frac{1}{\sigma_{e,q}^2}$ for the covariate model. The posterior distributions of $\boldsymbol{\psi_q}$ and $\sigma_{e,q}^2$ have exactly the same forms as the coefficients in substantive model. After imputing the missing covariates, the conditional posterior distributions of the substantive model parameters, the conditional posterior distribution of the regression coefficients in probit regression, and the

posterior distribution of the latent propensity $\boldsymbol{r}_{\boldsymbol{y}}^{*}$, and the posterior predictive distribution of $y_{i(miss)}$ are the same as the ones when the covariates are complete (please see the previous section). More details are presented in the supplemental materials.

When the covariates have a missing completely at random (MCAR) mechanism where the probability of missingness of the covariates is unrelated to either observed or unobserved variables, we still can use the illustrated methodology in this section to impute missing covariates and missing outcome, and estimate the substantive model.

As mentioned previously, the previous approach readily accommodates incomplete binary, ordinal, and nominal covariates with MAR missingness mechanisms, as does our later extension for NMAR covariates. We just need to extend the previous equations to incorporate a cumulative probit model for ordinal variables or a multinomial probit model for nominal responses (e.g., Agresti, 2018; J. H. Albert & Chib, 1993; Johnson & Albert, 2006; McCulloch & Rossi, 1994). Take a binary covariate as an example, where we can use a binary probit regression to model the incomplete responses. In this scenario, the model introduces an underlying normally distributed random variable for an incomplete binary covariate, with the variance of the latent covariate is usually fixed at 1 for identification (this is the same probit model used for MNAR missingness on the outcome). A threshold divides the distribution of the latent continuous covariate into two segments, such that the latent continuous covariate is below the threshold when the binary variable equals zero and above the threshold when the binary variable equals one. Compared to the continuous covariates, instead of specifying Equation (14) for incomplete binary covariate, we specify Equation (14) for each latent continuous covariate and conditional on all other latent continuous covariates. Note that this latent continuous covariate is different from the aforementioned latent missingness propensity $\boldsymbol{r}^{*}$ when a variable is MNAR, although it does use the same probit regression framework. We refer the interested reader to Du et al. (Manuscript submitted for publication) and Enders et al. (Advance online publication).

Markov Chain Monte Carlo (MCMC) computational algorithm

We propose a Gibbs sampling algorithm to sample $\boldsymbol{\beta}$, σ_{ε}^2, $\boldsymbol{\gamma}$, r_{yi}^*, and $y_{i(miss)}$ from their aforementioned posterior distributions, and to obtain the posterior inferences based on the Monte Carlo samples (Gelfand & Smith, 1990). The Gibbs sampling algorithm is an iterative procedure that estimates the variables one at a time in a sequence (Gelfand & Smith, 1990): estimate regression coefficients while holding all other variables that their current values; estimate the residual variance while holding all other variables constant, and so forth. More specifically, it invokes the following steps: (a) estimate the substantive model regression coefficients and residual variance, (b) estimate the selection model's regression coefficients, (c) estimate the missingness propensity scores, (d) estimate outcome' missing values (e) estimate the covariate model's regression coefficients and residual variance, and (f) estimate covariates' missing values. Each of these steps treats the current values of all other unknowns as fixed constants. When the posterior distribution is not accessible or difficult to derive, the Metropolis-Hastings algorithm draws the posterior samples.

The MCMC algorithm gives a posterior distribution of each parameter, and we can use these quantities to conduct Bayesian inference for the substantive model parameters (i.e., $\boldsymbol{\beta}$ and σ_{ε}^2). Alternatively, one can save the imputations of missing data at regular intervals in the MCMC chain (e.g., save a data set every 1000 iterations) and use the filled-in data sets for a multiple imputation analysis (Rubin, 1987; Schafer, 1997). When frequentist estimation (e.g., ordinary least squares estimation) is applied to the imputed complete data to estimate the parameters in the substantive model, this leads to a hybrid procedure (Bayesian techniques are used for imputation and frequentist methods are used for parameter estimation).

The full cadre of step-by-step Gibbs sampler procedure is given below.

0. Initialization step: set initial values for $\boldsymbol{\beta}^{(0)}$, $\sigma_{\varepsilon}^{2(0)}$, $\boldsymbol{\gamma}^{(0)}$, and $\boldsymbol{r}_y^{*(0)}$, and $y_{i(miss)}^{(0)}$ (for the individuals who have missing outcome). For the individuals who have missing covariates, set initial values

for $x_q^{(0)}$, $\boldsymbol{\psi_q}^{(0)}$, and $\sigma_{e,q}^{2(0)}$.

1. In the t^{th} iteration, given covariates in the substantive model ($\boldsymbol{X}$), the imputed outcomes in the previous iteration ($\boldsymbol{y}^{(t-1)}$), and the residual variance of the substantive model in the previous iteration ($\sigma_{\varepsilon}^{2(t-1)}$), sample $\boldsymbol{\beta}^{(\mathrm{t})}$ from Equation (8).

2. Given $\boldsymbol{X}$, $\boldsymbol{y}^{(t-1)}$, and $\boldsymbol{\beta}^{(\mathrm{t})}$, sample $\sigma_{\varepsilon}^{2(t)}$ from Equation (9).

3. Given $\boldsymbol{y}^{(t-1)}$, the predictors $\boldsymbol{M}$ in the missingness model, and $\boldsymbol{r_y^*}^{(\mathrm{t}-1)}$, sample $\boldsymbol{\gamma}^{(\mathrm{t})}$ from Equation (10).

4. For all individuals, given $\boldsymbol{r}$, $\boldsymbol{y}^{(t-1)}$ and $\boldsymbol{\gamma}^{(\mathrm{t})}$, sample $\boldsymbol{r}_y^{*(\mathrm{t})}$ from Equation (11).

5. For the individual i who has missing outcome (i.e., $r_i = 1$), given the covariates and/or auxiliary variables, $\boldsymbol{\beta}^{(\mathrm{t})}$, $\sigma_{\varepsilon}^{2(\mathrm{t})}$, $\boldsymbol{\gamma}^{(\mathrm{t})}$, and $r_{yi}^{*(\mathrm{t})}$, sample $y_{i(miss)}^{(t)}$ from Equation (12). Repeat step 5 for all individuals who have missing outcome and obtain a set of updated imputed outcomes, $\boldsymbol{y}^{(t)}$.

6. For the qth incomplete predictor, given $\boldsymbol{X_{-q}}^{(t-1)}$ and $\sigma_{e,q}^{2(t-1)}$, sample $\boldsymbol{\psi_q}^{(t)}$ from Equation (2) in the supplemental materials.

7. For the qth incomplete predictor, given $\boldsymbol{X_{-q}}^{(t-1)}$ and $\boldsymbol{\psi_q}^{(t)}$, sample $\sigma_{e,q}^{2(t)}$ from Equation (3) in the supplemental materials.

8. For the individual i who has the qth missing covariate, given $y_i^{(t)}$, $\boldsymbol{X_{-qi}}^{(t-1)}$, $\boldsymbol{\beta}^{(\mathrm{t})}$, $\sigma_{\epsilon}^{2(t)}$, $\boldsymbol{\psi_q}^{(t)}$, and $\sigma_{e,q}^{2(t)}$, sample $x_{qi(miss)}^{(t)}$ from Equation (13) by Metropolis-Hastings algorithm. Repeat steps 6 to 8 for all individuals who have missing covariates and impute all covariates. The missing interaction terms can be calculated by the updated components. For example, $(x_{1i}x_{2i})^{(t)} = x_{1i}^{(t)} \times x_{2i}^{(t)}$.

9. Repeat steps 1 to 8 until the MCMC chains reach convergence and provide sufficient posterior samples.

Bayesian Estimation of BLVSM with MNAR Covariates

In this section, we further extend the previous ideas by allowing missingness of an incomplete covariate to depend on the unobserved covariate variable itself. Additionally, the missing covariate can depend on the unobserved scores of other covariates, observed covariates, auxiliary variables, and the outcome. The missingness of the outcome may also be conditional on the unobserved outcome scores or the unobserved covariate scores, or it can be MAR (or even MCAR). As mentioned previously, the literature and existing methods of MNAR generally have focused on MNAR outcomes, except a few studies investigating MNAR covariates (Huang et al., 2005; Ibrahim et al., 1999, 2005). The existing literature has worked on MNAR covariates, MNAR outcome, MAR covariates, or MAR outcome. Our approach is more general than the previous models because it can accommodate MNAR covariates, MAR covariates, MAR/MNAR outcome, or all of them simultaneously. Although putting the models in the previous literature together also can accommodate all of the aforementioned cases, this paper is the first one which systematically presents all cases. Additionally, there are three major differences between our work and the work from Ibrahim's group. First, Ibrahim et al. (1999) and Ibrahim et al. (2005) proposed algorithms to handle MNAR covariates in the expectation–maximization (EM) framework, whereas we use Bayesian statistics. As a Bayesian method, our method can use informative priors and incorporate prior information (e.g., from existing papers or pilot results). Second, Ibrahim et al. (1999) and Ibrahim et al. (2005) focused on the sequential specification which we present in the supplemental material, whereas we focus on the separate or fully conditional specification which may be more widely used when researchers assume a linear relationship between covariates. Third, our separate specification assumes that the missingness latent variables (propensities) are independent after controlling for the influence of the cause of missingness, whereas Ibrahim's sequential specification assumes that the missingness latent variables are still correlated after controlling for the influence of the cause of missingness. However, Ibrahim's sequential specification may cause nonconvergence and we may need to simplify the model by assuming

that the missingness latent variables are independent after controlling for the influence of cause of missingness, which is kind of back to the separate specification. We refer audiences to the supplemental material for more details about the sequential specification.

Suppose there are Q partially observed covariates which are MNAR, and $K - Q$ fully observed covariates. Similar to the case where the outcome is MNAR. A missing indicator $r_{x,q}$ is used to indicate the missingness of the qth missing covariate x_q. $r_{x,q}$ does not apply to the interaction terms in the substantive model. An underlying random variable $r^*_{x,q}$ captures the latent propensity of missingness for the qth missing covariate x_q. When we assume the missingness of x_q is conditional on the unobserved x_q and the other covariates, the missingness or selection model is

$$\boldsymbol{r^*_{x,q}} = \boldsymbol{X}\boldsymbol{\gamma_{x,q}} + \boldsymbol{\zeta_{x,q}}, \tag{17}$$

where $\boldsymbol{r^*_{x,q}}$ is a $N \times 1$ vector of latent missingness propensities of x_q for N individuals, $\boldsymbol{X} = (\boldsymbol{1}, \boldsymbol{x_q}, \boldsymbol{X_{-q}})$ is a $N \times (K + 1)$ matrix and $\boldsymbol{X_{-q}}$ are the covariates other than x_q, $\boldsymbol{\gamma_{x,q}}$ is a $(K + 1) \times 1$ vector, and $\boldsymbol{\zeta_{x,q}}$ is a $N \times 1$ vector following $N(\boldsymbol{0}, \boldsymbol{I})$ (as mentioned above, the residual variance is fixed at one for identification). The missing indicator $r_{x,q}$ is conditional on the propensity $r^*_{x,q}$ through a probit regression, $P(r_{x,q} = 1|\boldsymbol{X}) = \Phi(\boldsymbol{X}\boldsymbol{\gamma_{x,q}})$. This is the same model as before. The $r^*_{x,q}$ variable is a latent missingness variable scaled as a z-score, and the right side of the expression features potential predictors of missingness (typically, x_q plus other substantive model variables).

Besides the missingness probit model of $r_{x,q}$, a regressive covariate model specifies the relation between x_q and all other covariates, $p\left(\boldsymbol{x_q}|\boldsymbol{X_{-q}}, \boldsymbol{\psi_q}, \sigma^2_{e,q}\right)$, which is the same as Equation (14) if the separate specification is used. Thus, the joint conditional distribution of x_q and its latent missingness variable $r^*_{x,q}$ is factored into three components: the substantive model, the covariate model of x_q, and the

missingness model of x_q,

$$p\left(\boldsymbol{x_q}, \boldsymbol{r^*_{x,q}}|\boldsymbol{y}, \boldsymbol{X_{-q}}, \boldsymbol{\beta}, \sigma^2_\varepsilon, \boldsymbol{\psi_q}, \sigma^2_{e,q}, \boldsymbol{\gamma_{x,q}}\right) \propto p\left(\boldsymbol{y}|\boldsymbol{x_q}, \boldsymbol{X_{-q}}, \boldsymbol{\beta}, \sigma^2_\varepsilon\right) p\left(\boldsymbol{x_q}|\boldsymbol{X_{-q}}, \boldsymbol{\psi_q}, \sigma^2_{e,q}\right) p\left(\boldsymbol{r^*_{x,q}}|\boldsymbol{\gamma_{x,q}}, \boldsymbol{X}\right). \tag{18}$$

If one prefers to use the sequential specification for the covariate model and the missingness model, the details are provided in the supplemental materials. With the sequential specification, we can accommodate the nonlinear relations between the covariates and latent missingness variables.

When interaction or curvilinear terms in the substantive model are incomplete, each missing covariate appears multiple times in the substantive model and we need to extract it from all the relevant components. We use the example where there is one partially observed covariate x_1 and one partially observed interaction term. The substantive model, the missingness probit model for y, the regressive covariate model for x_1, and the missingness probit model for x_1 are respectively

$$\begin{aligned}
y_i &= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + \varepsilon_i \quad \varepsilon_i \sim N\left(0, \sigma^2_\varepsilon\right) \\
r^*_{yi} &= \gamma_0 + \gamma_1 y + \zeta_i \quad e_{i1} \sim N\left(0, 1\right) \\
x_{1i} &= \psi_{1,0} + \psi_{1,1} x_{2i} + e_{i1} \quad e_{i1} \sim N\left(0, \sigma^2_{e,1}\right) \\
r^*_{x,1i} &= \gamma_{x,1,0} + \gamma_{x,1,1} x_{1i} + \zeta_{x,1i} \quad \zeta_{x,1i} \sim N\left(0, 1\right).
\end{aligned} \tag{19}$$

x_1 appears in both x_1 and $x_1 x_2$. Based on Equation (18), for individual i, the conditional posterior

distribution of x_{1i} is

$$p\left(x_{1i}|\cdot\right) \propto p\left(y_i|x_{1i}, x_{2i}, \boldsymbol{\beta}, \sigma_{\varepsilon}^2\right) p\left(x_{1i}|x_{2i}, \boldsymbol{\psi_1}, \sigma_{e,1}^2\right) p\left(r_{x,1i}^*|x_{1i}, x_{2i}, \boldsymbol{\gamma_{x,1}}\right) \tag{20}$$
$$= N\left(\frac{\sigma_{e,1}^2\left(\beta_1+\beta_3 x_{2i}\right)\left(y_i-\beta_0-\beta_2 x_{2i}\right)+\sigma_{\varepsilon}^2\left(\psi_{1,0}+\psi_{1,1} x_{2i}\right)+\sigma_{e,1}^2\sigma_{\varepsilon}^2\gamma_{x1,1}\left(r_{x,1i}^*-\gamma_{x1,0}\right)}{\left(\beta_1+\beta_3 x_{2i}\right)^2\sigma_{e,1}^2+\sigma_{\varepsilon}^2+\gamma_{x1,1}^2\sigma_{e,1}^2\sigma_{\varepsilon}^2},\right.$$
$$\left.\frac{\sigma_{e,1}^2\sigma_{\varepsilon}^2}{\left(\beta_1+\beta_3 x_{2i}\right)^2\sigma_{e,1}^2+\sigma_{\varepsilon}^2+\gamma_{x1,1}^2\sigma_{e,1}^2\sigma_{\varepsilon}^2}\right).$$

The analytical form for imputing x_{1i} is complex, and it is specific to the particular substantive model, Equation (19). Generally, the Metropolis-Hastings algorithm is suggested to draw the posterior samples in practice, as this algorithm allows the procedure to extend to covariate sets with an arbitrary composition and general missing data patterns. Nevertheless, the equation follows the same basic form as before. That is, the mean and variance combines information from three regressions – the substantive model, the covariate model, and the selection model in which x_1 plays a role. If x_1 plays a role in multiple selection models, we need to consider all of them.

The posterior predictive distribution of $r_{x,qi}^*$ is conditional on the missing predictor indicator $r_{x,qi}$ and imputed predictors $\boldsymbol{X_i}$ ($\boldsymbol{X_i} = (1, x_{1i})$ in the example of Equation (19)),

$$p\left(r_{x,qi}^*|\cdot\right) = \begin{cases} N\left(\boldsymbol{X_i}\boldsymbol{\gamma_{x,q}}, 1\right) I\left(\kappa = 0, \infty\right) & r_{x,qi} = 1 \\ N\left(\boldsymbol{X_i}\boldsymbol{\gamma_{x,q}}, 1\right) I\left(-\infty, \kappa = 0\right) & r_{x,qi} = 0 \end{cases}. \tag{21}$$

In words, this equation says that latent missingness scores are drawn from one of two normal distributions, both of which are centered at the predicted z-score from the regression equation (the mean of the normal distribution) and have a fixed variance equal to 1. Specifically, if x_{qi} is observed, a latent score should be drawn from the region of the normal curve below 0 (the fixed threshold parameter), as this area corresponds to the region occupied by indicator scores of $r_{x,qi} = 0$. Otherwise, if x_{qi} is missing, a latent score should

be drawn from the region of the normal curve above the threshold, as this area corresponds to the region occupied by indicator scores of $r_{x,qi} = 1$.

We employ weakly informative prior $p(\gamma_{x,q}) = N(0, b)$ for all coefficients in $\boldsymbol{\gamma_{x,q}}$ in the missingness model of x_q to facilitate convergence in MNAR case, as for $\boldsymbol{\gamma}$ in the missingness model of y. Similar to Equation (10), the conditional posterior distribution of $\boldsymbol{\gamma_{x,q}}$ is a multivariate normal distribution,

$$p(\boldsymbol{\gamma_{x,q}}|\cdot) = MN\left(\hat{\boldsymbol{\gamma_{x,q}}} = \Sigma_1^{-1}\boldsymbol{X'r^*_{x,q}}, \Sigma_1 = \left(\frac{1}{b} \times I + \boldsymbol{X'X}\right)^{-1}\right). \tag{22}$$

$\boldsymbol{X} = (\mathbf{1}, \boldsymbol{x_1})$ in the example of Equation (19). In words, Equation (22) says that the selection model's regression coefficients are drawn from a multivariate normal distribution, and the center and variance are determined by the latent data and current imputed data. The conditional posterior distribution of $\sigma^2_{e,q}$ is the same as Equation (3) in the supplemental materials, and the conditional posterior distribution of $\boldsymbol{\psi_q}$ is the same as Equation (2) in the supplemental materials. After imputing the missing covariates, the conditional posterior distributions of the substantive model parameters, the conditional posterior distribution of the regression coefficients in probit regression for y, and the posterior distribution of the latent propensity $\boldsymbol{r^*_y}$ for y, and the posterior predictive distribution of $y_{i(miss)}$ are the same as the ones when the covariates are MAR/complete.

When a categorical covariate has a MNAR mechanism, we will need two probit regression models (one for the binary covariate and another for its missingness model), and we will need both the latent covariate and the latent missingness propensity. For example, suppose x_1 is a incomplete binary variable with a MNAR mechanism. First, x_1^* is the latent x_1. x_1^* does not appear in the substantive analysis but only is used to in the covariate model to impute the missing x_1. The first probit regression model describes the distribution of x_1^* . A threshold (usually fixed at 0) divides the normal distribution of x_1^* into two segments, such that the x_1^* is below the threshold if $x_1 = 0$ and above the threshold if $x_1 = 1$. Second, the latent

missingness propensity $r^*_{x,1}$ provides a latent missingness propensity for x^*_1. The second probit regression model is a missingness model, which captures how x^*_1 influence the missingness propensity of $r^*_{x,1}$. This missingness model is the same as Equation (17) with a difference that the propensity is on the latent covariate and the predictors in the probit model are also on the latent variable scales. For example, $r^*_{x,1} = \gamma_{x,1,0} + \gamma_{x,1,1}x^*_1 + \zeta_{x,1}$. We will illustrate more details in Simulation Study S2 in the supplemental materials.

Markov Chain Monte Carlo (MCMC) computational algorithm

The step-by-step Gibbs sampler procedure when the outcome is MNAR and the covariates are MNAR is given below. The first five steps generate estimates for the substantive analysis model, the missingness model, and missing outcomes. Steps 6-7 generate estimates for the parameters in the covariate model. Steps 8-10 target on the missingness model for covariates and missing covariates.

0. Initialization step: set initial values for $\boldsymbol{\beta}^{(0)}$, $\sigma_{\varepsilon}^{2(0)}$, $\boldsymbol{\gamma}^{(0)}$, and $\boldsymbol{r}_y^{*(0)}$, and $y^{(0)}_{i(miss)}$ (for the individuals who have missing outcome). For the individuals who have missing predictors, set initial values for $x_q^{(0)}$, $\boldsymbol{\psi_q}^{(0)}$, $\sigma_{e,q}^{2(0)}$, $\boldsymbol{\gamma_{x,q}}^{(0)}$, and $\boldsymbol{r^*_{x,q}}^{(0)}$,

1-7. Steps 1-7 are exactly the same as the ones in the section of MAR covariates.

8. For the qth incomplete predictor, given $\boldsymbol{X_{-q}}^{(t-1)}$ and $\boldsymbol{r^*_{x,q}}^{(t-1)}$, sample $\boldsymbol{\gamma_{x,q}}^{(t)}$ from Equation (22).

9. For all individuals, given $\boldsymbol{X}^{(t-1)}$ and $\boldsymbol{\gamma_{x,q}}^{(t)}$, sample $\boldsymbol{r^*_{x,q}}^{(t)}$ from Equation (21).

10. For the individual i who has the qth missing predictor, sample $x^{(t)}_{qi(miss)}$ from Equation (18) by Metropolis-Hastings algorithm. Repeat steps 6 to 10 for all individuals who have missing covariates and impute all covariates.

11. Repeat steps 1 to 10 until the MCMC chains reach convergence and provide sufficient posterior samples.

Simulation Study 1: MAR Covariates

Simulation Study 1: Simulation Design

This simulation study examines the performance of the proposed Bayesian latent variable selection model, BLVSM, when x_1 is MAR and y is MNAR. We also conducted a simulation where y is MNAR and covariates are complete (see Simulation Study S1 in the supplemental materials). The substantive model for the simulation is the widely-used moderated regression model in Equation (4). The missing values on the outcome y were generated as a function of y itself and the missing values on x_1 were generated as a function of x_2. We varied the values of the following four factors. (1) The first factor is the sample size (SZ= 50, 100, 200, 500, or 1000). (2) The second factor is the missing data proportion/probability for y (P_y = 0.1, 0.2, or 0.4). (3) The third factor is the pseudo coefficient of determination between the cause of missingness and the latent propensities of y, $R^2_{r^*_y}$ = 0.1, 0.25, or 0.5 (McKelvey & Zavoina, 1975). When $R^2_{r^*_y}$ is large, the MNAR selection process of y is strong and the missingness of y heavily depends on y. When $R^2_{r^*_y}$ is 0, the missingness of y is independent from y, which leads to a missing completely at random (MCAR) case whereby the probability of missingness is not related to any observed variables or unobserved variables. Because in practice, we don't know the true missingness mechanism, it is important to check whether estimating a model for the missingness negatively impacts the substantive analysis when the MNAR selection process is very weak and almost MCAR. (4) The fourth factor is the missing data proportion for x_1 (P_{x1} = 0.1, 0.2, or 0.4). (5) The fifth factor is the pseudo coefficient of determination between x_2 and the latent propensities of x_1 ($R^2_{r^*_{x1}}$ = 0.1, 0.25, or 0.5). When $R^2_{r^*_{x1}}$ is large, the MAR selection process of x_1 is strong. When $R^2_{r^*_{x1}}$ is 0, the missingness of x_1 does not depend on x_2 or any observed/unobserved scores, which is a MCAR case. The coefficient of determination in the substantive model R^2_y (the proportion of the variance in the outcome that is predictable from the covariates, x_1, x_2, and x_1x_2) is fixed at 0.13, a medium effect size (Cohen, 1988). By fixing the mean and variance of y at $E(y) = 5$ and $var(y) = 10$, fixing the means of x_1 and x_2 at 0, fixing the regression coefficients at 1, and

fixing the correlation between x_1 and x_2 at 0.3, we can solve the variances of x_1 and x_2 and the residual variance σ_ε^2 given a specific value of R_y^2 (the mean and variance of the interaction term is determined by formulas in Bohrnstedt & Goldberger (1969)).

We used a probit regression equation to link missingness probabilities of y to the values of y ($r_{yi}^* = \gamma_0 + \gamma_1 y_i + \zeta_i$ with $\zeta_i \sim N(0,1)$) and another probit regression equation to link missingness probabilities of x_1 to the values of x_2 ($r_{x,1i}^* = \gamma_{x,1,0} + \gamma_{x,1,1} x_{2i} + \zeta_{x,1i}, \zeta_{x,1i} \sim N(0,1)$). Using a latent variable formulation for probit regression (Agresti, 2018; Johnson & Albert, 2006), we derived γ_0, γ_1, $\gamma_{x,1,0}$, and $\gamma_{x,1,1}$ that produced the desired missing data proportion P_y, R_y^2, P_{x1} and $R_{r_{x1}^*}^2$ values. Finally, we sampled a missing data indicator for each observation (0 = observed, 1 = missing) from a binomial distribution with success rate equal to that observation's missingness probability from the probit regression model, and we deleted scores with indicator values of one. But when we estimate the missing covariates, we do not need to estimate its probit model. Instead, we specify a covariate model $x_{1i} = \psi_0 + \psi_1 x_{2i} + e_i, e_i \sim N\left(0, \sigma_e^2\right)$ and use Equation (13) to impute the missing x_1. There were 1000 replications under each condition.

As a comparison, we first applied the ordinary least squares estimation (OLS) to the original complete data. The results from the complete data are treated as the simulation baselines. We also applied a misspecified Bayesian method with assuming that both x_1 and y are MAR to the incomplete data. When we assume the outcome is MAR, we simply draw y from $N\left(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i}, \sigma_\varepsilon^2\right)$ and ignore the missingness model. In addition to the Bayesian summaries of the model parameters (we refer it to as the *full Bayesian approach*), we also saved imputed data sets and applied multiple imputation inference. More specifically, we used our proposed approach to impute missing data, saved 20 complete sets of data from the posterior samples of the converged chains, and conducted multiple imputation (20 imputations were suggested by Graham et al. (2007)). We used the OLS estimator to fit Equation (4) to the multiply imputed data sets, and pooled estimates and standard errors based on Rubin's pooling rules (see

Rubin (1976) and Schafer (1997) for more details).

The following priors of the model parameters were used: $p(\beta) \propto 1$, $p\left(\sigma_{\epsilon}^{2}\right) = IG(1,1)$, $p(\psi) \propto 1$, $p\left(\sigma_{e,q}^{2}\right) \propto \frac{1}{\sigma_{e,q}^{2}}$, and $p(\gamma) = N(0, var = 10)$. The priors on β and ψ are the Jeffreys prior, which is widely known as noninformative because it gives whole parameter space an equal prior probability. The priors on γ and σ_{ϵ}^{2} are all weakly informative priors which contain little information but facilitate convergence. $IG(1,1)$ is a flat distribution, which gives an almost equal prior probability in a relatively wide parameter space, and $N(0, var = 10)$ has a relatively large prior variance.[2] The initial burn-in period was 10^4, after that we checked convergence every 2×10^4 iterations, and all the iterations before the converged 2×10^4 iterations were treated as the final burn-in period. Geweke (1992) convergence diagnostic was used. If after 20 times, the chain still did not converge, we claimed nonconvergence. The simulation was coded in R.

We compared the performance of BLVSM with that of the misspecified method with an MAR assumption, based on the accuracy of the point estimates via the bias and relative bias, and both the accuracy and precision via the coverage rates of the 95% confidence intervals (CI) or posterior credible intervals for each parameter of interest. Denote a parameter of interest by θ. The bias and relative bias are calculated by averaging $\hat{\theta}_r - \theta$ and $\frac{\hat{\theta}_r - \theta}{\theta} \times 100\%$ (when $\theta \neq 0$) respectively across the 1000 replications, where $\hat{\theta}_r$ is the point estimate from the r^{th} replication. $\hat{\theta}_r$ was calculated using the posterior mean or mode. We consider average relative bias (averaging over 1000 replications) lower than 10% as ignorable (L. K. Muthén & Muthén, 2002). The OLS estimation from the complete-data (pre-deletion) is used as a reference. Both the quantile-based probability (QBP) interval and the highest posterior density (HPD) interval were obtained as credible intervals. The coverage rate was calculated as the proportion of the 95% credible/confidence intervals covering the true parameter value. We considered coverage rates between 91% and 98% as satisfactory (L. K. Muthén & Muthén, 2002).

Simulation Study 1: Simulation Results

The convergence rates in all conditions were over 94% for BLVSM, and they were 100% for the misspecified method with an MAR assumption. This finding is practically important because modeling an MNAR process is computationally challenging relative to an MAR analysis. The fact that the more complicated modeling task reduced convergence rates by only 6% across a wide range of conditions is encouraging. The detailed summaries of each method under each condition are found in the supplemental materials. The main findings on the performance of each method are summarized below.

We first focus on BLVSM. In terms of biases, when the sample size exceeded 200 ($SZ > 200$), the biases were negligible. When the sample size was less than or equal to 200 ($SZ \leq 200$), the posterior mode – the most likely value for a parameter from its posterior distribution – was found to be less biased than the posterior mean for σ^2_ε, γ_0, and γ_1. The posterior mode and mean were similar for β_0, β_1, β_2, and β_3 which are the main focus of the substantive model. The point estimates from multiple imputation were very close to the posterior means for the full Bayesian approach. As such, we focus on the the posterior mode from the full Bayesian approach. We found that the influence of the sample size (SZ), the missingness proportion (P_y), the pseudo coefficient of determination between y and the latent propensities of y ($R^2_{r^*_y}$), the missingness proportion of x_1 (P_{x1}), and the pseudo coefficient of determination between the x_2 and the latent propensities of x_1 ($R^2_{r^*_{x1}}$) were consistent for β_0, β_1, β_2, and β_3. In the interest of space, we select the minimum and maximum values of P_y, $R^2_{r^*_y}$, P_{x1}, and $R^2_{r^*_{x1}}$ to illustrate the bias results in figures. More specifically, the average relative biases of β_3 are presented in Figure 1, the average relative biases of β_0 and β_1 are presented in Figure 2, and the average relative biases of β_2 and σ^2_ε are presented in Figure 3. In figures, each cell represents a combination of different levels of P_y, $R^2_{r^*_y}$, P_{x1}, and $R^2_{r^*_{x1}}$, and different lines represent the relative biases from BLVSM with an MNAR process, the misspecified model with an MAR assumption, and the complete data respectively. The row panel effects reflect the influence from P_{x1} and $R^2_{r^*_{x1}}$, and the column panel effects reflect the influence from P_y and $R^2_{r^*_y}$. When the sample

size was greater than or equal to 200, the relative biases of β_0, β_1, β_2, β_3, and σ_ε^2 were ignorable and very close to the OLS estimates from the complete-data. We found that when the sample size was less than or equal to 100, BLVSM provided smaller biases for β_0, β_1, β_2, and β_3 when (1) the missingness proportion of y was smaller (see cells across column panels with different P_y in Figures 1-3), (2) the MNAR selection process of y was weaker (see cells across column panels with different $R^2_{r^*_y}$ in Figures 1-3), (3) the missingness proportion of x_1 was smaller (see cells across row panels with different P_{x1} in Figures 1-3), and (4) the MAR selection process of $x1$ was weaker (see cells across row panels with different $R^2_{r^*_{x1}}$ a in Figures 1-3). But the positive biases of the σ_ε^2 estimates were smaller with a larger $R^2_{r^*_y}$ and a larger P_{x1} (see Figure 3). When the MNAR selection process was weak, incorporating the latent missingness model in BLVSM did not negatively impact the substantive analysis and again provided unbiased estimates.

Next, consider the misspecified method with an MAR assumption. In terms of biases, when the MNAR selection process was strong ($R^2_{r^*_y} \geq 0.25$), the estimates of β_0, β_1, β_2, β_3, and σ_ε^2 in the misspecified method with an MAR assumption were underestimated relative to their true values (i.e., > 10% relative bias) and increasing sample size did not effectively improve the point estimates. Only when the MNAR selection process was weak ($R^2_{r^*_y} = 0.1$) could the misspecified model with an MAR assumption approximate the complete-data estimates; the MAR-based analysis was still inferior to BLVSM in this case, although the difference was not practically significant (see Figures 1-3).

[Figures 1-3]

In terms of the coverage rates, the QBP intervals had slightly better coverage rates than the HPD intervals in both BLVSM and the misspecified Bayesian method with an MAR assumption. The differences between the confidence intervals from multiple imputation and the QBP intervals in the full Bayesian approach were trivial. As such, we focus on the QBP intervals in the full Bayesian approach. The coverage rates for β_0, β_1, β_2, and σ_ε^2 are presented in Figure 4 after fixing $R^2_{r^*_{x1}} = 0.5$ and $P_{x1} = 0.4$ (the most severe MAR case for the incomplete covariate) and selecting the minimum and maximum values of P_y and

$R^2_{r^*_y}$ (i.e., the missingness proportion of y and strength of the selection mechanism, respectively). The effects of P_y, $R^2_{r^*_y}$, and the sample size are the column panel effect, row panel effect, and the x-axis effect within each cell, respectively. For BLVSM, the coverage rates of β_0, β_1, β_2, and σ^2_ε were close to the nominal level (95%) except when the sample size was less than or equal to 100 (see the x-axis effect in Figure 4). With a small sample size, there could be undercoverage. For the misspecified method with an MAR assumption, the coverage rates for β_0, β_1, β_2, and σ^2_ε were far below the nominal level and even close to 0 in some cases (see Figure 4). Even when the MNAR selection process was weak ($R^2_{r^*_y} = 0.1$), severe undercoverage was observed.

[Figure 4]

Turning to the covariate model, the estimates of ψ_0, ψ_1, and σ^2_e were unbiased and coverage rates were acceptable from BLVSM across all the manipulated conditions (see the supplemental materials). For these particular parameters, the misspecified method with an MAR assumption provided similar but slightly worse estimates than the correctly-specified model, BLVSM. When the sample size was greater than or equal to 50 ($SZ \geq 50$), the estimated ψ_0, ψ_1, and σ^2_e from the MNAR and MAR methods were essentially the same as the OLS estimates from the complete-data. Besides the parameters in the substantive and covariate models, BLVSM estimates γ_0 and γ_1 in the missingness model for y. The estimation of these coefficients was challenging, and we observed negative biases and undercoverage of credible intervals when the sample size was not large enough ($SZ \leq 200$, see the supplemental materials). This was an interesting and encouraging finding, given that the substantive model parameters were largely unaffected by the biases in the missingness model or at least achieved their optimal properties at a smaller sample size.

Simulation Study 2: MNAR Covariates

Simulation Study 2: Simulation Design

The moderated regression model in Equation (4) again served as the substantive model for the simulation. This simulation study examined the performance of the proposed method when both x_1 and y are MNAR. The missing values on the outcome y were generated as a function of y itself and the missing values on x_1 were generated as a function of x_1 itself. The conditions were the same as the second simulation (SZ, P_y, $R^2_{r^*_y}$, and P_{x1}) except that the fifth factor is the pseudo coefficient of determination between x_1 and the latent propensities of x_1 which reflects the strength of MNAR selection process of x_1 ($R^2_{r^*_{x1}}$ = 0.1, 0.25, or 0.5). We used a probit regression equation to link missingness probabilities of y to the values of y ($r^*_{yi} = \gamma_0 + \gamma_1 y_i + \zeta_i$ with $\zeta_i \sim N(0,1)$) and another probit regression equation to link missingness probabilities of x_1 to the values of x_1 ($r^*_{x,1i} = \gamma_{x,1,0} + \gamma_{x,1,1} x_{1i} + \zeta_{x,1i}, \zeta_{x,1i} \sim N(0,1)$). When we estimate the missing x_1, we not only estimate the probit model but also estimate the covariate model $x_{1i} = \psi_0 + \psi_1 x_{2i} + e_i, e_i \sim N\left(0, \sigma^2_e\right)$. The coefficient of determination in the substantive model R^2_y is fixed at 0.13.

As a comparison, we applied the OLS estimator to the original complete data and applied the MAR method with assuming that both x_1 and y are MAR assumption to the incomplete data. When we assume x_1 is MAR, we draw x_1 from Equation (13). In addition, we conducted multiple imputation. The following priors of the model parameters were used: $p(\beta) \propto 1$, $p\left(\sigma^2_\epsilon\right) = IG(1,1)$, $p(\psi) \propto 1$, $p\left(\sigma^2_{e,q}\right) \propto \frac{1}{\sigma^2_{e,q}}$, $p(\gamma) = N(0,10)$ and $p(\gamma_{x,q}) = N(0,10)$ (weakly informative prior to facilitate convergence). The burn-in period is the same as Study 1.

Simulation Study 2: Simulation Results

The convergence rates of all conditions were again over 94% when modeling an MNAR process on both the outcome and the explanatory variable. Given the complexity of this modeling problem, we found

this result very encouraging. The detailed summaries of each method under each condition are in the supplemental materials, and the main findings are summarized below.

We again focus on the posterior mode from the full Bayesian approach. We select the minimum and maximum values of P_y, $R^2_{r^*_y}$, P_{x1} and the pseudo coefficient of determination between x_1 and the latent missingness propensities of x_1 (i.e., $R^2_{r^*_{x1}}$) to illustrate the bias results (Figures 5-7). In figures, each cell represents a combination of different levels of P_y, $R^2_{r^*_y}$, P_{x1}, and $R^2_{r^*_{x1}}$, and different lines represent the relative biases from BLVSM with an MNAR process, the misspecified model with an MAR assumption, and the complete data respectively. The row panel effects reflect the influence from P_{x1} and $R^2_{r^*_{x1}}$, and the column panel effects reflect the influence from P_y and $R^2_{r^*_y}$. We found that when the sample size was greater than or equal to 200, the relative biases of β_0, β_1, β_2, β_3, and σ^2_ε were ignorable and very close to the OLS estimates from the complete-data. When the sample size was less than or equal to 100, relative biases for β_0, β_1, β_2, and β_3 in the Bayesian latent variable approach decreased as (1) the missingness proportions of x_1 and y decreased (see cells with different P_y and P_{x1} in Figures 5-7), (2) the MNAR selection process of y became weaker (see cells across column panels with different $R^2_{r^*_y}$ in Figures 5-7), and (3) the MNAR selection process of x_1 became weaker (see cells across row panels with different $R^2_{r^*_{x1}}$ in Figures 5-7). But a larger $R^2_{r^*_y}$ decreased the positive biases of the σ^2_ε estimate (see Figure 7). In contrast, the misspecified method with an MAR assumption had the pattern of underestimating β_0, β_1, β_2, β_3, and σ^2_ε unless when the MNAR selection processes of the outcome and x_1 were weak ($R^2_{r^*_y} = R^2_{r^*_{x1}} = 0.1$).

[Figures 5-7]

Turning to coverage rates, the influence from P_{x1} and $R^2_{r^*_{x1}}$ on the QBP coverage rates for β_0, β_1, β_2, and σ^2_ε was not large. Therefore, the coverage rates for β_0, β_1, β_2, and σ^2_ε are presented in Figure 8 for the $R^2_{r^*_{x1}} = 0.5$ and $P_{x1} = 0.4$ (the severest missingness case) conditions along with the minimum and maximum values of $R^2_{r^*_y}$ and P_y. Similar to Simulation Study 1, the coverage rates of β_0, β_1, β_2, and σ^2_ε from BLVSM were close to the nominal level (95%) except when the sample size was small (e.g.,

$SZ = 50$ or 100; see Figure 8). The coverage rates for β_0, β_1, β_2, and σ_ε^2 from the misspecified method with an MAR assumption were again too low (see Figure 8).

[Figure 8]

In addition to the substantive model, we examined the covariate model and missingness model parameters. The estimates of ψ_0, ψ_1, and σ_e^2 in the covariate model were unbiased, and coverage rates were acceptable from BLVSM across most conditions, except when the sample size was small (see the supplemental materials). When the sample size was greater than or equal to 100, the estimates of ψ_0, ψ_1, and σ_e^2 from BLVSM were essentially the same as the OLS estimates from the complete-data. However, different from the Study 1 where the covariate is MAR, in the current simulation study, the misspecified method with an MAR assumption produced biased point estimates and considerable undercoverage relative to the correctly specified model even with a large sample size, particularly when the MNAR selection process was strong (e.g., a large value of $R^2_{r^*_{x1}}$). Finally, considering the missingness model parameters in BLVSM, γ_0, γ_1, $\gamma_{x,1,0}$, and $\gamma_{x,1,1}$, we observed biases and undercoverage when the sample size was less than or equal to 500 (see the supplemental materials), but this bias apparently had no material impact on the substantive analysis. Again, it is encouraging that bias was relegated to a part of the analysis that is not of substantive interest.

As mentioned previously, BLVSM readily extends to accommodate categorical covariates as well. We conducted an additional simulation to generalize the result of Simulation Study 2 to binary covariates. This extra simulation study considered the same scenario as Simulation Study 2, but with x_1 as a binary variable. The performance of BLVSM was similar to that in the case of continuous covariates. In the interest of space, we refer interested readers to Simulation Study 2S in the supplemental materials for additional details.

Simulation Study 3: Misspecification

The challenging part of using MNAR methods is that we cannot identify or prove which selection model or missingness mechanism are the true ones. Ibrahim et al. (2005) summarized two different views on specifying the selection model. First, one can let data empirically determine the selection model by comparing the model fit index. One can then use the likelihood ratio or AIC to evaluate the fit of each model. However, "it is often the case that little information is contained in the data regarding alternative nonignorable models" (page 341). Alternatively, one can view a set of MNAR analyses with different selection models as a sensitivity analysis that examines how stable substantive model parameter estimates are across different missingness models. Our method and related software provide the opportunity to conduct sensitivity analysis with various selection models. The previous simulations clearly show that failing to model an MNAR process (e.g., by fitting an MAR analysis to data where the true process is MNAR) is detrimental. Thus, the practical danger for researchers is specifying a model with too few predictors of missingness, as it will usually be difficult to know which covariates to include in a given selection model. One potential remedy for this model specification problem is to deploy rich models that include all variables in the selection model. The question is whether misspecifying the selection model in this way has a detrimental effect on the substantive model parameters.

To provide some practical guidelines, we conducted additional simulations that examined the impact of misspecifying missingness models by including too many (or too few) predictors. Ibrahim et al. (2005) cautioned against making the selection model too complex and suggested that the main-effects model usually is adequate. Thus, our simulation only focused on the main-effects model (i.e., the missingness model included all variables in the substantive analysis but did not include interaction effects). Past literature on selection models suggests that including predictors from the analysis model may induce collinearity problems that are detrimental to estimation (for details refer to Puhani, 2000; Stolzenberg & Relles, 1990, 1997). The simulations in this section suggest that this finding does not extend to BLVSM,

and we ultimately recommend including all variables from the substantive model in the missingness model.

Simulation Study 3: Simulation Design

The moderated regression model in Equation (4) again served as the substantive model for the simulation. We considered three types of missingness scenarios. First, x_1 was missing due to y, which indicated a MAR scenario. In this scenario, we considered three selection models: $r^*_{x,1i} = \gamma_{x,1,0} + \gamma_{x,1,1}x_{1i} + \zeta_{x,1i}$ (misspecified), $r^*_{x,1i} = \gamma_{x,1,0} + \gamma_{x,1,1}y_i + \zeta_{x,1i}$ (correctly specified), and $r^*_{x,1i} = \gamma_{x,1,0} + \gamma_{x,1,1}x_{1i} + \gamma_{x,1,2}x_{2i} + \gamma_{x,1,3}y_i + \zeta_{x,1i}$ (over-specified). Second, x_1 was missing completely at random (MCAR). In this scenario, we considered two over-specified selection models: $r^*_{x,1i} = \gamma_{x,1,0} + \gamma_{x,1,1}x_{1i} + \zeta_{x,1i}$ and $r^*_{x,1i} = \gamma_{x,1,0} + \gamma_{x,1,1}x_{1i} + \gamma_{x,1,2}x_{2i} + \gamma_{x,1,3}y_i + \zeta_{x,1i}$. The first two scenarios investigate the situation where a researcher incorrectly applies a selection model to an analysis where the missingness model is unnecessary. In the third scenario, x_1 was missing due to both x_1 and y, which indicated a mixture of MAR and MNAR processes. In this scenario, we considered three selection models: $r^*_{x,1i} = \gamma_{x,1,0} + \gamma_{x,1,1}x_{1i} + \zeta_{x,1i}$ (misspecified), $r^*_{x,1i} = \gamma_{x,1,0} + \gamma_{x,1,1}x_{1i} + \gamma_{x,1,2}y_i + \zeta_{x,1i}$ (correctly specified), and $r^*_{x,1i} = \gamma_{x,1,0} + \gamma_{x,1,1}x_{1i} + \gamma_{x,1,2}x_{2i} + \gamma_{x,1,3}y_i + \zeta_{x,1i}$ (over-specified). We also conducted additional simulations that considered similarly misspecifications of the outcome's missingness model, but the results were similar to that for missingness on the covariate. Therefore, we focus on these three representative simulations. The simulation conditions were as follows. The sample size (SZ) varied as 200, 500, and 1000. The coefficient of determination in the substantive model R^2_y was fixed at 0.13. The missing data proportion for x_1 (P_{x1}) was 0.2. and the pseudo coefficient of determination between the cause of missingness and the latent propensities of x_1 ($R^2_{r^*_{x1}}$) was 0.25.

Simulation Study 3: Simulation Results

The convergence rates of all conditions and scenarios were 100%. We present the relative biases of posterior mode estimates (when the true value is 0, absolute biases are presented instead) and the coverage

rates of QBP intervals in Table 1. Specifically, we focus on the estimates of substantive models and selection models when the selection models are over-specified. For example, in the first scenario, when the selection model is specified as $r^*_{x,1i} = \gamma_{x,1,0} + \gamma_{x,1,1}x_{1i} + \gamma_{x,1,2}x_{2i} + \gamma_{x,1,3}y_i + \zeta_{x,1i}$, the true values of $\gamma_{x,1,1}$ and $\gamma_{x,1,2}$ are 0.

When x_1 was missing due to y (Scenario 1, an MAR process), estimates were biased if we misspecified the fitted selection model by omitting the true cause of missingness (i.e., $r^*_{x,1i} = \gamma_{x,1,0} + \gamma_{x,1,1}x_{1i} + \zeta_{x,1i}$). Regardless of how much we increased the sample size, the biases in the substantive model estimates did not decrease, and the coverage rates actually got worse. When the selection model was correctly specified in the sense that it included only the true cause of missingness (i.e., $r^*_{x,1i} = \gamma_{x,1,0} + \gamma_{x,1,1}y_i + \zeta_{x,1i}$), the bias values and coverage rates were within the acceptable range, even with a sample size as small as 200. When the selection model was over-specified by including all variables from the substantive analysis model as predictors (i.e., $r^*_{x,1i} = \gamma_{x,1,0} + \gamma_{x,1,1}x_{1i} + \gamma_{x,1,2}x_{2i} + \gamma_{x,1,3}y_i + \zeta_{x,1i}$), performance was generally quite good. Although the biases of some parameters (in both the substantive model and the selection model) were relatively large with a sample size of 200, the coverage rates of the parameters of the substantive model were otherwise acceptable. Additionally, when the sample size increased, the biases of the parameters in both the substantive model and the selection model decreased and were near zero. Specifically, the estimates of $\gamma_{x,1,1}$ and $\gamma_{x,1,2}$ were very close to the true value 0. That is, x_1 and x_2 should have no influence on the missingness of x_1.

Next, consider the situation where x_1 was missing completely at random (Scenario 2). In this case, the fitted selection models (i.e., $r^*_{x,1i} = \gamma_{x,1,0} + \gamma_{x,1,1}x_{1i} + \zeta_{x,1i}$ and $r^*_{x,1i} = \gamma_{x,1,0} + \gamma_{x,1,1}x_{1i} + \gamma_{x,1,2}x_{2i} + \gamma_{x,1,3}y_i + \zeta_{x,1i}$) are over-specified because there are no causes of missingness. Although the biases of some parameters (in both the substantive model and the selection model) were relatively large with a sample size of 200, but the biases and coverage rates of the parameters

of the substantive model were otherwise acceptable. Similar to the Scenario 1, when the sample size increased, the biases of the parameters in both the substantive model and the selection model decreased and approximated zero. Specifically, the estimates of $\gamma_{x,1,1}$, $\gamma_{x,1,2}$, and $\gamma_{x,1,3}$ were very close to the true value 0. That is, the two covariates and outcome should have no influence on the missingness of x_1.

Finally, consider the scenario where x_1 was missing due to both x_1 and y (Scenario 3). If we omitted one of the true causes of missingness in the selection model (i.e., $r^*_{x,1i} = \gamma_{x,1,0} + \gamma_{x,1,1}x_{1i} + \zeta_{x,1i}$), the biases of estimates and coverage rates of the parameters in the substantive model were unacceptable, regardless of the sample size. When the selection model is correctly specified ($r^*_{x,1i} = \gamma_{x,1,0} + \gamma_{x,1,1}x_{1i} + \gamma_{x,1,2}y_i + \zeta_{x,1i}$), the parameter estimates and coverage rates were acceptable even with a sample size of 200. Finally, when the selection model was over-specified by including unnecessary predictors (i.e., $r^*_{x,1i} = \gamma_{x,1,0} + \gamma_{x,1,1}x_{1i} + \gamma_{x,1,2}x_{2i} + \gamma_{x,1,3}y_i + \zeta_{x,1i}$), the biases of the parameters in both the substantive model and the selection model decreased as the sample size increased. The biases of the substantive model and selection model parameters reached acceptable levels (e.g., < 10% relative bias) at a sample size of 500 and 1000, respectively. Consistent with the findings from the previous two simulation studies, the substantive model parameter estimates were acceptable even when the missingness model parameter estimates were biased.

[Table 1]

The simulations investigating misspecifications provide the following conclusions: (1) omitting the true cause of missingness caused biases and disrupted coverage rates, (2) correct specification yielded accurate estimates and acceptable coverage rates even with a relatively small sample size (i.e., 200), and (3) adding extra, unnecessary predictors to the missingness part of the selection model caused biases when the sample size was relatively small, but the coverage rates were close to the nominal level. In the current simulation, as the sample size increased to 500, the biases of substantive model parameter estimates due to over-specification generally diminished to below the 10% threshold. We would like to highlight that when

the missingness model is over-specified, the true parameters of the unnecessary predictors are 0 although in samples they are never estimated to be exactly 0. Simulation Study 3 shows that the estimates from the over-specified model can have ignorable biases and acceptable coverage rates. Our conclusions seem to offer a fairly clear prescription for researchers applying these models: specify selection models that are more inclusive, including all variables in the analysis model. This strategy provides a realistic possibility of obtaining approximately unbiased parameter estimates in sample sizes that are typical of the behavior sciences, whereas adopting a more restrictive specification that may omit potential predictors of missingness risks inducing substantial biases. We illustrate this approach in the ensuing real data analysis example.

A Real Data Example

In a marital satisfaction study at the University of California, Los Angeles, a sample of 431 first-married couples were asked to rate their marriage on a 8-item scale twice in 2012 and 2014 respectively. The sum of the ratings was treated as an index of marital satisfaction. We are interested in whether wives' marital satisfaction at the first wave $(WS1)$ had an influence on the husbands' marital satisfaction at the second wave $(HS2)$ after controlling husbands' marital satisfaction at the first wave $(HS1)$, husbands' education levels (EDU), and husbands' stress levels (STR). Therefore, the substantive model is

$$HS2 = \beta_0 + \beta_1 HS1 + \beta_2 WS1 + \beta_3 EDU + \beta_4 STR + \varepsilon, \tag{23}$$
$$\varepsilon \sim N\left(0, \sigma_\varepsilon^2\right)$$

The missing proportions in the husbands' marital satisfaction scores at the two waves and the wives' marital satisfaction scores at the second wave were 13.2%, 22.3%, and 13.0%, respectively. The education and stress level scores of husbands were complete.

All MAR- and MNAR-based methods including BLVSM rely heavily on untestable assumptions of missingness. We cannot prove the missingness is MNAR or MAR. Similarly, we cannot prove whether a specific MNAR selection model is appropriate for a given data set. Given the inherent uncertainty associated with conducting NMAR analysis, we followed recommendations from Ibrahim et al. (2002) by conducting sensitivity analysis that apply different assumptions of missingness to the same data. Additionally, following our conclusions from Simulation Study 3, we modeled NMAR processes with rich selection models that included all variables in the substantive analyses (i.e., $HS1$, $WS1$, $HS2$, EDU, and STR) as the predictors. When the MCMC chains had difficulty in converging, we simplified the selection models by removing predictor variables. In this real data example, we considered eight assumptions (Tables 1 and 2). We used the forthcoming Blimp 3 application (Keller et al., 2019) to apply BLVSM, and the Blimp code (both separate and sequential specifications) is illustrated in the Appendix. A typical application might consist of the substantive regression, an selection model for outcome's missingness that features all variables from the substantive model, and a selection model for a covariate's missingness, again with all variables from the analysis as predictors. As mentioned previously, Blimp allows for binary, ordinal, or nominal covariates (and categorical outcomes), and it readily accommodates interactive or non-linear effects. Depending on the assumptions, it is possible to fit these models in other packages. For example, specialized Bayesian programs like WinBugs or JAGS could certainly estimate these models. Based on the R technical manual and Lüdtke et al. (2020), the R package 'mdmb' (Robitzsch and Luedtke, 2019) can handle an MAR or MNAR outcome and covariates. The 'mdmb' package uses the sequential approach. Bayesian estimation can be used in conjunction with a structural equation modeling framework such as Mplus (L. Muthén & Muthén, 1998–2017) to incorporate selection models for an outcome or a covariate, with two caveats (even when users specify syntax by themselves). First, incomplete covariates are assumed to be normally distributed, although outcomes can be binary and ordinal. Second, because the SEM framework is grounded in the multivariate normality assumption, interactive or non-linear effects

with missing data would be estimated with potentially substantial biases (e.g., Bartlett et al., 2015; Enders et al., 2018; Erler et al., 2016; Kim et al., 2015; Seaman et al., 2012; Van Buuren et al., 2006).

In the first analysis, we assumed the covariates ($HS1$ and $WS1$) and the outcome ($HS2$) were missing at random (MAR). We used five MCMC chains with different starting values, and the Gelman–Rubin diagnostic (Gelman & Rubin, 1992) was used to check the convergence of the five chains after the burning period (Gelman & Rubin, 1992). As mentioned in the simulation studies, we can get Bayesian estimates of β_0, β_1, β_2, β_3, β_4 and σ_ε^2 from the MCMC algorithm in the full Bayesian approach, and we can also generate multiple imputations. We report the posterior mode estimates, the posterior standard deviations (i.e., standard deviation from posterior samples; SD), the quantile-based probability (QBP) credible intervals, and the deviance information criterion (DIC) in Table 1. We reject $H_0 : \beta = 0$ when the QBP interval does not cover 0. In addition to the full Bayesian inferences, we applied multiple imputation with 100 imputed sets of data from the posterior samples by BLVSM (each chain provided 20 sets of data). The R and Mplus code for pooling the estimates and standard errors are illustrated in the Appendix. The point estimates, standard errors, the Akaike information criterion (AIC), and the Bayesian information criterion (BIC) are in Table 2. Both the full Bayesian and multiple imputation results showed that wives' marital satisfaction at the first wave $(WS1)$ could significantly predict husbands' marital satisfaction at the second wave $(HS2)$ after controlling for other variables, and $\hat{\beta}_2$ was about 0.2. Additionally, the husbands' marital satisfaction at the first wave $(HS1)$ and husbands' education levels (EDU) could significantly predict husbands' marital satisfaction at the second wave.

Although our simulation results suggest that over-parameterizing a selection model by incorporating an inclusive set of covariates is not problematic at current sample size, specifying complex selection models may not be feasible in every dataset. We reduce model complexity by removing predictor variables in the selection models if the complex selection models fail to converge because the data contain insufficient information to estimate such a complex model. Additionally, we would not recommend treating

all variables are MNAR, as it seems unlikely that such models would converge in practice. Rather, we suggest a model-building procedure that researchers assume one variable is MNAR first (when such a process is theoretically justified), and move to the analyses where two and more variables are MNAR. This is the process we applied here.

In the second analysis, we assumed that $WS1$ and $HS2$ were MAR and $HS1$ was MNAR, where the missingness of $HS1$ depended on $HS1$, $WS1$, $HS2$, EDU, and STR. Using both the full Bayesian approach and multiple imputation based on BLVSM, the estimated coefficients and posterior standard deviations/multiple imputation errors of the substantive model are in Table 2. In addition, we provide the estimates of the probit missingness model for $HS1$ (e.g., $\gamma_{HS1|HS1}$ and$\gamma_{HS1|HS2}$) in Table 2. We compared the substantive parameter estimates in the current analysis to the ones in the first analysis without any selection model. As a practical guide, we investigated how much the estimates changed in posterior standard deviation units (SD; we used the SDs in the first analysis, which were similar in magnitude to the imputation-based standard errors). We found the estimates and hypothesis testing results of the substantive model did not noticeably differ from those with only MAR assumptions (the first analysis), and largest change (e.g., in $\hat{\beta}_2$) was equivalent to about 0.7 SDs.

Third, we assumed that $HS1$ and $HS2$ were MAR, and $WS1$ was MNAR, with the missingness of $WS1$ depended on $HS1$, $WS1$, $HS2$, EDU, and STR. Compared to the first analysis, $\hat{\beta}_1$ changed about 1.4 posterior SDs ($\hat{\beta}_1$ was 0.636 and 0.558 in the first and third assumptions, respectively) and other estimates changed less than 1 posterior SD (see Table 2).

Fourth, we assumed that $HS1$ and $WS1$ were MAR, and $HS2$ was MNAR. In this analysis, different from the previous analyses, wives' marital satisfaction at the first wave failed to predict husbands' marital satisfaction at the second wave (β_2). Compared to the first analysis, most of the estimates changed more than 1 SD. More specifically, $\hat{\sigma}^2_\varepsilon$ changed about 6.8 SDs, $\hat{\beta}_1$ changed about 1.6 SDs ($\hat{\beta}_1$ was 0.636 and 0.548 in the first and fourth assumptions, respectively), and $\hat{\beta}_2$ changed about 1.5 SDs ($\hat{\beta}_2$ was 0.214

and 0.13 in the first and fourth assumptions, respectively; see Table 2). We tried to explore why the estimates changed dramatically. We removed the predictors in the selection model one by one and found when $HS1$ was removed, the estimates and inferences were similar to those in the first analysis.

After pairing the substantive analysis with one selection model, we next fitted models that incorporated a pair of selection models. Then in the fifth model, we assumed that $HS1$ and $WS1$ were MNAR and $HS2$ was MAR. Including all variables as predictors in the missingness model led to nonconvergence, therefore we did not include the outcome, $HS2$, in the selection models. There is no clear guideline of how to simplify the selection model. We suggest removing one variable at a time, The estimates of the substantive model and selection models are in Table 3. This model had the smallest DIC, AIC, and BIC. The estimates and hypothesis testing results of the substantive model did not noticeably differ from those in the first analysis with only MAR assumptions. The largest change is that $\hat{\beta}_4$ changed about 0.2 SDs (see Table 3 for estimates).

In the sixth model, we assumed that $HS1$ and $HS2$ were MNAR, and $WS1$ was MAR. Compared to the first analysis, β_2 was not significant anymore and most of the estimates changed more than 1 SD (e.g., $\hat{\sigma}^2_\varepsilon$ changed about 8.1 SDs, $\hat{\beta}_0$ changed about 3.4 SDs, and $\hat{\beta}_2$ changed about 1.8 SDs; see Table 3 for estimates). The change of the estimates probably is probably due to assuming $HS2$ was MNAR based on the results in the fourth assumption.

In the seventh models, we assumed that $WS1$ and $HS2$ were MNAR, and $HS1$ was MAR. However, in the seventh model, including all variables as predictors in the missingness model led to nonconvergence, therefore we did not include the outcome, $HS2$, in the missingness model of $WS1$. Compared to the first analysis, β_2 was not significant anymore and most of the estimates changed more than 1 SD (e.g., $\hat{\sigma}^2_\varepsilon$ changed about 6 SDs and $\hat{\beta}_2$ changed about 2.3 SDs; see Table 3 for estimates). Again, we think the change of the estimates is due to assuming $HS2$ was MNAR.

Finally, in the eighth model, we assumed $HS1$, $WS1$ and $HS2$ were all MNAR. However, the five

chains did not converge, even when we only include one predictor in the missingness model and used 6×10^4 iterations. Because the posterior distributions of multiple parameters converged to two different modes, we did not pursue this model. We suggest that researchers should be cautious when assuming more than one covariate are MNAR, paying careful attention to convergence diagnostics such as the Gelman–Rubin diagnostic statistic.

Considered as a whole, when we assumed $HS2$ was MNAR (the fourth, sixth, and seventh assumptions), the influence of wives' marital satisfaction at the first wave on husbands' marital satisfaction at the second wave (β_2) was no longer significant, which was different from the result obtained when $HS2$ was MAR. We also found when excluding $HS1$ from predictors in the selection model of $HS2$ in the model 3, the results were more consistent with the other models. The difficulty with the discrepancy between the models is that a researcher cannot verify which model is more plausible based on the data. If one's substantive knowledge suggests that the NMAR process might be plausible, then a reasonable course of action is to present multiple sets of results (e.g., including the full sensitivity analysis in an online supplement). We don't necessarily view such discrepancies across models as insurmountable or inherently problematic, nor do we feel that it is necessary for a researcher to choose one set of results - in fact, there is little basis for such a choice beyond one's expert opinion about the plausibility of different processes. Online supplemental documents offer researchers unlimited space with which to report multiple sets of results, and we find it just fine to declare that different assumptions about the missingness process led to somewhat different conclusions for certain model parameters. This won't always be the case, but sometimes it will. Certainly, reporting two sets of results is a better alternative that choosing just one, particularly when that choice involves effects that are significant under one assumption and non-significant under another. We believe the importance of this exercise stems from doing a thorough job of trying to understand if, how, and why one's analysis results are sensitive to missing data assumptions, not choosing "the" best model.

[Tables 2-3]

Conclusion

In real data analysis, usually a missing at random mechanism (MAR; missingness is related to observed data but not to the unobserved values of itself) or missing completely at random mechanism (MCAR; missingness is unrelated to either observed or unobserved data) is assumed. However, it is possible that the underlying missingness mechanism is MNAR. If we ignore the possibility of an MNAR selection process by inappropriately applying an MAR-based procedure, previous research and our own simulations have shown that parameter estimates generally were biased (Enders, 2011; Fitzmaurice et al., 2012; Graham, 2009; Yang & Maxwell, 2014). Building on one of the major MNAR modeling frameworks – the selection model – this paper outlined a Bayesian latent variable selection model, BLVSM, that accommodates an MNAR process on the outcome, covariates, or both. This procedure offers a number of compelling advantages: it (a) has a strong theoretical foundation in the Bayesian framework, (b) can be either applied in a full Bayesian framework where parameters in the substantive model are calculated in MCMC steps, or in a multiple imputation framework where the missing data are imputed by MCMC steps and the parameters in the substantive model are estimated by frequentist methods later, (c) easily handles complete or incomplete covariates (due to MCAR, MAR, or MNAR), (d) allows the incomplete MAR or MNAR covariates to involve interactions, non-linear terms, and random slopes, and (e) accommodates categorical variables. The procedure is implemented in a forthcoming release of the software package Blimp (Keller et al., 2019). We are unaware of other packages that offer these modeling possibilities, although the R package 'mdmb' can estimate some selection models in the EM framework.

Computer simulation results suggest that BLVSM is quite effective when the outcome is MNAR and the covariates are complete, MAR, or MNAR (regardless of whether covariates are continuous or binary). Except when the sample size was small (e.g., $SZ \leq 100$), estimates tracked closely with those from a

complete-data analysis. More specifically, the substantive model parameter estimates were unbiased and their coverage rates were acceptable, even when parameters of the missingness model exhibited bias (these parameters required large samples to achieve their optimal properties). Moreover, convergence failures were rare, even when simultaneously modeling an MNAR process for the outcome and covariate. In addition, we found that the following factors influenced the performance of BLVSM: the sample size, the coefficient of determination in the substantive model, the missingness proportion of y, the strength of MNAR selection process of y, the missingness proportion of covariates, and the strength of MAR/MNAR selection process of covariates. With a relatively small sample size, when the outcome was less predictable from the covariates, the missingness proportions of the covariates and the outcome were larger, and the missingness process of the covariates and the outcome were more MNAR and/or MAR, the performance of BLVSM was less satisfactory. When the sample size was large, the factors barely influenced the performance. Multiple imputation as a hybrid approach provided similar results as the full Bayesian method, thus researchers have various options for applying our approach. As noted in the introduction, the literature has largely focussed on MNAR processes for the outcome variable except some work investigating MNAR covariates (Huang et al., 2005; Ibrahim et al., 1999, 2005). Our approach is quite flexible because it can accommodate MNAR covariates, MAR covariates, MAR/MNAR outcome, or all of them simultaneously. Although putting the models in the previous literature together also can accommodate all of the aforementioned cases, this paper is the first one which systematically presents all cases. Additionally, our work and the work from Ibrahim's group have differences in terms of estimation method and assumptions.

We explored the robustness of the proposed method in Simulation Study 3. Based on the results, we suggest specifying an inclusive selection model for each variable. When MCMC chains have difficulty in converging, we can simplify the selection models to make the computation process easier. Although we suggest an inclusive selection model, in practice, it is not feasible to include all variables as predictors in

the selection model because it may lead to nonconvergence. Based on prior knowledge and existing theories, researchers can select several important predictors to enter the selection model. Based on our simulation results, there is no need to force the selection model to have only one or two predictors. We also suggest conducting sensitivity analysis to investigate how much the results change across missingness assumptions and missingness models. More specifically, we suggest a model-building procedure. We begin with assuming all variables are MAR. Then we assume one variable is MNAR (when such a process is theoretically justified), and move to the analyses where two and more variables are MNAR, as we illustrated in the real data example.

Due to the scope and word limitation of this paper, we only focus on the missingness patterns that can be handled by selection models. BLVSM has not generalized to other missingness patterns such as pattern mixture models yet.

In sum, our paper outlined a new Bayesian latent variable selection model for an MNAR process. When missingness is truly MNAR, computer simulations suggest that the proposed model can offer substantial improvement over methods that apply an incorrect MAR assumption. The Blimp application offers a user-friendly environment for implementing BLVSM.

References

Agresti, A. (2018). *An introduction to categorical data analysis*. Hoboken, NJ: JohnWiley & Sons. doi: 10.1002/0470114754

Aiken, L. S., West, S. G., & Reno, R. R. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks, CA: Sage. doi: 10.1016/0886-1633(93)90008-d

Albert, J. H., & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, *88*(422), 669–679. doi: 10.1080/01621459.1993.10476321

Albert, P. S., & Follmann, D. A. (2000). Modeling repeated count data subject to informative dropout. *Biometrics*, *56*(3), 667–677. doi: 10.1111/j.0006-341x.2000.00667.x

Bartlett, J. W., Seaman, S. R., White, I. R., Carpenter, J. R., & Initiative*, A. D. N. (2015). Multiple imputation of covariates by fully conditional specification: accommodating the substantive model. *Statistical Methods in Medical Research*, *24*(4), 462–487. doi: 10.1177/0962280214521348

Beunckens, C., Molenberghs, G., Verbeke, G., & Mallinckrodt, C. (2008). A latent-class mixture model for incomplete longitudinal gaussian data. *Biometrics*, *64*(1), 96–105. doi: 10.1111/j.1541-0420.2007.00837.x

Bohrnstedt, G. W., & Goldberger, A. S. (1969). On the exact covariance of products of random variables. *Journal of the American Statistical Association*, *64*(328), 1439–1442. doi: 10.1080/01621459.1969.10501069

Cohen, J. (1988). *Statistical power analysis for the social sciences*. Hillsdale, NJ: Erlbaum. doi: 10.4324/9780203771587

Cowles, M. K. (1996). Accelerating monte carlo markov chain convergence for cumulative-link generalized linear models. *Statistics and Computing*, *6*(2), 101–111. doi: 10.1007/bf00162520

Daniels, M. J., & Hogan, J. W. (2008). *Missing data in longitudinal studies: Strategies for bayesian modeling and sensitivity analysis*. Boca Raton, FL: Chapman and Hall/CRC. doi: 10.1201/9781420011180

Dantan, E., Proust-Lima, C., Letenneur, L., & Jacqmin-Gadda, H. (2008). Pattern mixture models and latent class models for the analysis of multivariate longitudinal data with informative dropouts. *The International Journal of Biostatistics*, *4*(1), 1–26. doi: 10.2202/1557-4679.1088

Demirtas, H., & Schafer, J. L. (2003). On the performance of random-coefficient pattern-mixture models for non-ignorable drop-out. *Statistics in Medicine*, *22*(16), 2553–2575. doi: 10.1002/sim.1475

Diggle, P., & Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *43*(1), 49–73. doi: 10.2307/2986113

Du, H., Mena, S., & Alacam, E. (Manuscript submitted for publication). Compatibility in multiple imputation with single incomplete predictor and multiple incomplete predictors.

Enders, C. K. (2011). Missing not at random models for latent growth curve analyses. *Psychological Methods*, *16*(1), 1. doi: 10.1037/a0022640

Enders, C. K. (in press). *Applied missing data analysis 2.0*. Guilford Press.

Enders, C. K., Du, H., & Keller, B. T. (Advance online publication). A model-based imputation procedure for multilevel regression models with random coefficients, interaction effects, and non-linear terms. *Psychological Methods*. doi: 10.1037/met0000228

Enders, C. K., Hayes, T., & Du, H. (2018). A comparison of multilevel imputation schemes for random coefficient models: Fully conditional specification and joint model imputation with random covariance matrices. *Multivariate Behavioral Research*, *53*(5), 695–713. doi: 10.1080/00273171.2018.1477040

Erler, N. S., Rizopoulos, D., Jaddoe, V. W., Franco, O. H., & Lesaffre, E. M. (2019). Bayesian imputation of time-varying covariates in linear mixed models. *Statistical Methods in Medical Research*, *28*(2), 555–568. doi: 10.1177/0962280217730851

Erler, N. S., Rizopoulos, D., Rosmalen, J. v., Jaddoe, V. W., Franco, O. H., & Lesaffre, E. M. (2016). Dealing with missing covariates in epidemiologic studies: a comparison between multiple imputation and a full bayesian approach. *Statistics in Medicine*, *35*(17), 2955–2974. doi: 10.1002/sim.6944

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2012). *Applied longitudinal analysis* (Vol. 998). Hoboken, NJ: John Wiley & Sons. doi: 10.1201/9781420011579

Follmann, D., & Wu, M. (1995). An approximate generalized linear model with random effects for informative missing data. *Biometrics*, 151–168. doi: 10.2307/2533322

Foster, E. M., Fang, G. Y., & Group, C. P. P. R. (2004). Alternative methods for handling attrition: An illustration using data from the fast track evaluation. *Evaluation Review*, *28*(5), 434–464. doi: 10.1177/0193841x04264662

Galimard, J.-E., Chevret, S., Curis, E., & Resche-Rigon, M. (2018). Heckman imputation models for binary or continuous mnar outcomes and mar predictors. *BMC medical research methodology*, *18*(1), 90. doi: 10.1186/s12874-018-0547-1

Galimard, J.-E., Chevret, S., Protopopescu, C., & Resche-Rigon, M. (2016). A multiple imputation approach for mnar mechanisms compatible with heckman's model. *Statistics in Medicine*, *35*(17), 2907–2920. doi: 10.1002/sim.6902

Gelfand, A. E., & Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, *85*(410), 398–409. doi: 10.21236/ada208388

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472. doi: 10.1214/ss/1177011136

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *In bayesian statistics* (pp. 169–193). University Press. doi: 10.21034/sr.148

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1996). Introducing markov chain monte carlo. In W. R. Gilks, S. Richardson, & D. J. Spiegelhalter (Eds.), *Markov chain monte carlo in practice* (pp. 339 – 357). London: Chapman & Hall. doi: 10.1201/b14835

Glynn, R. J., Laird, N. M., & Rubin, D. B. (1986). Selection modeling versus mixture modeling with nonignorable nonresponse. In H. Wainer (Ed.), *Drawing inferences from self-selected samples* (pp. 115–142). New York, NY: Springer. doi: 10.1007/978-1-4612-4976-4$_1$0

Goldstein, H., Carpenter, J. R., & Browne, W. J. (2014). Fitting multilevel multivariate models with missing data in responses and covariates that may include interactions and non-linear terms. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *177*(2), 553–564. doi: 10.1111/rssa.12022

Gottfredson, N. C., Bauer, D. J., & Baldwin, S. A. (2014). Modeling change in the presence of nonrandomly missing data: Evaluating a shared parameter mixture model. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(2), 196–209. doi: 10.1080/10705511.2014.882666

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*, 549–576. doi: 10.1146/annurev.psych.58.110405.085530

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? some practical clarifications of multiple imputation theory. *Prevention Science*, *8*(3), 206–213. doi: 10.1007/s11121-007-0070-9

Grund, S., Lüdtke, O., & Robitzsch, A. (2018). Multiple imputation of missing data for multilevel models: Simulations and recommendations. *Organizational Research Methods*, *21*(1), 111–149. doi: 10.1177/1094428117703686

Hafez, M. S., Moustaki, I., & Kuha, J. (2015). Analysis of multivariate longitudinal data subject to nonrandom dropout. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(2), 193–201. doi: 10.1080/10705511.2014.936086

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, *57*(1), 97–109. doi: 10.1093/biomet/57.1.97

Heckman, J. J. (1976). 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* (pp. 475–492).

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*(1), 153–161. doi: 10.2307/1912352

Hedeker, D., & Gibbons, R. D. (1997). Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods*, *2*(1), 64. doi: 10.1037/1082-989x.2.1.64

Huang, L., Chen, M.-H., & Ibrahim, J. G. (2005). Bayesian analysis for generalized linear models with nonignorably missing covariates. *Biometrics*, *61*(3), 767–780. doi: 10.1111/j.1541-0420.2005.00338.x

Ibrahim, J. G., Chen, M.-H., & Lipsitz, S. R. (2002). Bayesian methods for generalized linear models with covariates missing at random. *Canadian Journal of Statistics*, *30*(1), 55–78. doi: 10.2307/3315865

Ibrahim, J. G., Chen, M.-H., Lipsitz, S. R., & Herring, A. H. (2005). Missing-data methods for generalized linear models: A comparative review. *Journal of the American Statistical Association*, *100*(469), 332–346. doi: 10.1198/016214504000001844

Ibrahim, J. G., Lipsitz, S. R., & Chen, M.-H. (1999). Missing covariates in generalized linear models when the missing data mechanism is non-ignorable. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *61*(1), 173–190. doi: 10.1111/1467-9868.00170

Johnson, V. E., & Albert, J. H. (2006). *Ordinal data modeling*. New York: Springer Science & Business Media. doi: 10.4135/9781412986311.n9

Keller, B. T., Enders, C. K., & Du, H. (2019). Blimp software manual (version 2.1). *Los Angeles, CA*. Retrieved from `Available at http://www.appliedmissingdata.com/multilevel-imputation.html`

Kim, S., Belin, T. R., & Sugar, C. A. (2018). Multiple imputation with non-additively related variables: Joint-modeling and approximations. *Statistical Methods in Medical Research*, *27*(6), 1683–1694. doi: 10.1177/0962280216667763

Kim, S., Sugar, C. A., & Belin, T. R. (2015). Evaluating model-based imputation methods for missing covariates in regression models with interactions. *Statistics in Medicine*, *34*(11), 1876–1888. doi: 10.1002/sim.6435

Little, R. J. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, *88*(421), 125–134. doi: 10.2307/2290705

Little, R. J. (1994). A class of pattern-mixture models for normal incomplete data. *Biometrika*, *81*(3), 471–483. doi: 10.1093/biomet/81.3.471

Little, R. J., & Rubin, D. B. (2014). *Statistical analysis with missing data*. Hoboken, NJ: John Wiley & Sons. doi: 10.1002/9781119013563.ch2

Lüdtke, O., Robitzsch, A., & West, S. G. (2020). Regression models involving nonlinear effects with missing data: A sequential modeling approach using bayesian estimation. *Psychological Methods*, *25*(2), 157–181. doi: 10.1037/met0000233

Mason, A., Richardson, S., Plewis, I., & Best, N. (2012). Strategy for modelling non-random missing data mechanisms in observational studies using bayesian methods. *Journal of Official Statistics*, *28*, 279–302.

McCulloch, R., & Rossi, P. E. (1994). An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics*, *64*(1-2), 207–240. doi: 10.1016/0304-4076(94)90064-7

McKelvey, R. D., & Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology*, *4*(1), 103–120. doi: 10.1080/0022250x.1975.9989847

Muthén, B., Asparouhov, T., Hunter, A. M., & Leuchter, A. F. (2011). Growth modeling with nonignorable dropout: Alternative analyses of the STAR$*dantidepressanttrial$. *Psychological Methods*, *16*$(1), 17 - -33. doi : 10.1037/a0022634$

Muthén, L., & Muthén, B. (1998–2017). *Mplus userâs guide. 8th edition*. Los Angeles, CA: Author.

Muthén, L. K., & Muthén, B. O. (2002). How to use a monte carlo study to decide on sample size and determine power. *Structural Equation Modeling*, *9*(4), 599–620. doi: 10.1207/s15328007sem0904$_8$

Puhani, P. (2000). The heckman correction for sample selection and its critique. *Journal of Economic Surveys*, *14*(1), 53–68. doi: 10.1111/1467-6419.00104

Robitzsch, A., & Luedtke, O. (2019). mdmb: Model based treatment of missing data [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=mdmb` (R package version 1.3-18)

Roy, J. (2003). Modeling longitudinal data with nonignorable dropouts using a latent dropout class model. *Biometrics*, *59*(4), 829–836. doi: 10.1111/j.0006-341x.2003.00097.x

Roy, J., & Daniels, M. J. (2008). A general class of pattern mixture models for nonignorable dropout with many possible dropout times. *Biometrics*, *64*(2), 538–545. doi: 10.1111/j.1541-0420.2007.00884.x

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592. doi: 10.1093/biomet/63.3.581

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys* (Vol. 81). John Wiley & Sons. doi: 10.1002/9780470316696

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. CRC press. doi: 10.1201/9781439821862

Seaman, S. R., Bartlett, J. W., & White, I. R. (2012). Multiple imputation of missing covariates with non-linear effects and interactions: an evaluation of statistical methods. *BMC medical research methodology*, *12*(1), 46.

Stolzenberg, R. M., & Relles, D. A. (1990). Theory testing in a world of constrained research design: The significance of heckman's censored sampling bias correction for nonexperimental research. *Sociological Methods & Research*, *18*(4), 395–415. doi: 10.1177/0049124190018004001

Stolzenberg, R. M., & Relles, D. A. (1997). Tools for intuition about sample selection bias and its correction. *American Sociological Review*, 494–507. doi: 10.1177/0049124190018004001

Van Buuren, S., Brand, J. P., Groothuis-Oudshoorn, C. G., & Rubin, D. B. (2006). Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation*, *76*(12), 1049–1064. doi: 10.1080/10629360600810434

Wu, M. C., & Bailey, K. R. (1989). Estimation and comparison of changes in the presence of informative right censoring: conditional linear model. *Biometrics*, *45*(3), 939–955. doi: 10.2307/2531694

Wu, M. C., & Carroll, R. J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*, 175–188. doi: 10.2307/2531905

Yang, M., & Maxwell, S. E. (2014). Treatment effects in randomized longitudinal trials with different types of nonignorable dropout. *Psychological Methods*, *19*(2), 188. doi: 10.1037/a0033804

Yuan, Y., & Little, R. J. (2009). Mixed-effect hybrid models for longitudinal data with nonignorable dropout. *Biometrics*, *65*(2), 478–486. doi: 10.1111/j.1541-0420.2008.01102.x

Zhang, Q., & Wang, L. (2017). Moderation analysis with missing data in the predictors. *Psychological Methods*, *22*(4), 649. doi: 10.1037/met0000104

Appendix

Blimp code for BLVSM, and R and Mplus code for Multiple Imputation

```
###############################################################################
######## BLIMP CODE FOR HS1, WS1, AND HS2 ARE MAR (SEPARATE SPECIFICATION)
###############################################################################
DATA: inputdata.csv;
VARIABLES: HS1 WS1 EDU STR HS2;
ORDINAL: ;
NOMINAL: ;
MISSING: 999;
CLUSTERID: ;
MODEL: HS2 ~ HS1 WS1 EDU STR;
SEED: 90291;
BURN: 10000;
THIN: 10000; #Each imputed dataset is save every 10000 iteratives;
NIMPS: 100;
CHAINS: 5 processors 5;
OPTIONS: psr covariatemodel;
SAVE:
separate = imp*.dat; #Save 100 imputed datasets for Mplus
stacked = imps.dat; #Save 1 compiled imputed dataset for R


###############################################################################
######## BLIMP CODE FOR WS1 IS MAR, AND HS1 AND HS2 ARE MNAR (SEPARATE SPECIFICATION)
###############################################################################
DATA: inputdata.csv;
VARIABLES: HS1 WS1 EDU STR HS2;
ORDINAL: ;
NOMINAL: ;
MISSING: 999;
CLUSTERID: ;
MODEL:
```

```
HS2 ~ HS1 WS1 EDU STR;
HS1.missing ~ WS1 HS1 EDU STR;
HS2.missing ~ WS1 HS1 HS2 EDU STR;
SEED: 90291;
BURN: 15000;
THIN: 10000;
NIMPS: 100;
CHAINS: 5 processors 5;
OPTIONS: psr covariatemodel;
SAVE:
separate = imp*.dat;
stacked = imps.dat;

##############################################################################
######## BLIMP CODE FOR WS1 IS MAR, AND HS1 AND HS2 ARE MNAR (SEQUENTIAL SPECIFICATION)
##############################################################################
DATA: inputdata.csv;
VARIABLES: HS1 WS1 EDU STR HS2;
ORDINAL: ;
NOMINAL: ;
MISSING: 999;
CLUSTERID: ;
MODEL: HS2 ~ HS1 WS1 EDU STR;
#Squential covariate model
HS1 ~ WS1 EDU STR;
WS1 ~ EDU STR;
#Squential selection model
HS2.missing ~ WS1 HS1 HS2 EDU STR HS1.missing;
HS1.missing ~ WS1 HS1 HS2 EDU STR;
SEED: 90291;
BURN: 15000;
THIN: 10000;
NIMPS: 100;
CHAINS: 5 processors 5;
```

```
OPTIONS: psr covariatemodel;
SAVE:
separate = imp*.dat;
stacked = imps.dat;

##############################################################################
######## MPLUS CODE
######## 100 copies of data sets must be arranged as imp1.dat, imp2.dat,..., imp100.dat.
##############################################################################
DATA:
file = impnames.dat;
type = imputation;
VARIABLE:
names = HS1 WS1 EDU STR HS2;
usevariables = HS1 WS1 EDU STR HS2; #dshidhsid;
MODEL:
HS2 on HS1 WS1 EDU STR;
OUTPUT:
stdyx;

##############################################################################
######## R CODE
##############################################################################
# Required packages
library(mitml)
library(rstudioapi)

# set working directory to location of R script
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
impdata <- read.table(paste0(getwd(), "/imps.dat"))
names(impdata) <- c("imputation", "HS1", "WS1", "EDU", "STR", "HS2")

# analyze data and pool estimates
implist <- as.mitml.list(split(impdata, impdata$imputation))
```

```
analysis <- with(implist, lm(HS2 ~ HS1+WS1+EDU+STR))
estimates <- testEstimates(analysis, var.comp = T, df.com = NULL)
estimates
```

Footnotes

[1]The procedure is that multiplying the two components in Equation (6) and finding a normal distribution for $\boldsymbol{y}$ which has the same kernel as the product.

[2]Based on our pilot simulations, if we used noninformative prior for γ (i.e., $p(\gamma) \propto 1$), sometimes we could get converged results but sometimes not, which depended on the data. The default prior for coefficients in probit regression with missing data is $N(0, 5)$ in Mplus. We found that prior variance of 5, 10 or 15 did not yield observably different results, and it could ensure convergence results in almost all cases. In addition, r^* is scaled as a z-score, and we checked various probit regressions to capture the relation of y and r^* under different scenarios. We found that γ was not large across conditions. Therefore, we use the prior variance of 10 in the normal prior of γ, which is still quite large but small enough to induce additional information that facilitates convergence. In real data analysis, researchers can modify this weakly informative prior based on each specific data.

Table 1: Simulations of misspecification

		β_0	β_1	β_2	β_3	σ^2_ε	$\gamma_{x,1,1}$ (x_{1i})	$\gamma_{x,1,2}$ (x_{2i})	$\gamma_{x,1,3}$ (y_i)
Scenario 1		x_1 was missing due to y (MAR)							
		$r^*_{x,1i} = \gamma_{x,1,0} + \gamma_{x,1,1}x_{1i} + \zeta_{x,1i}$							
SZ=200	Biases	-0.048	0.409	-0.281	-0.317	-0.099			
	Coverage rates	0.800	0.680	0.871	0.888	0.853			
SZ=500	Biases	-0.049	0.440	-0.291	-0.278	-0.088			
	Coverage rates	0.594	0.310	0.730	0.827	0.759			
SZ=1000	Biases	-0.048	0.434	-0.291	-0.278	-0.084			
	Coverage rates	0.330	0.067	0.518	0.672	0.560			
		$r^*_{x,1i} = \gamma_{x,1,0} + \gamma_{x,1,1}y_i + \zeta_{x,1i}$							
SZ=200	Biases	0.002	-0.014	0.005	-0.066	-0.023			
	Coverage rates	0.944	0.958	0.951	0.943	0.944			
SZ=500	Biases	-0.001	0.016	-0.004	-0.014	-0.012			
	Coverage rates	0.941	0.952	0.971	0.946	0.966			
SZ=1000	Biases	0.000	-0.003	0.005	-0.009	-0.006			
	Coverage rates	0.950	0.953	0.959	0.942	0.951			
		$r^*_{x,1i} = \gamma_{x,1,0} + \gamma_{x,1,1}x_{1i} + \gamma_{x,1,2}x_{2i} + \gamma_{x,1,3}y_i + \zeta_{x,1i}$							
SZ=200	Biases	-0.011	-0.124	0.018	-0.191	-0.031	(0.096)	(0.008)	0.118
	Coverage rates	0.955	0.931	0.949	0.927	0.935	0.894	0.945	0.945
SZ=500	Biases	-0.007	-0.049	0.016	-0.089	-0.014	(-0.009)	(0.009)	0.053
	Coverage rates	0.951	0.941	0.966	0.936	0.963	0.927	0.945	0.945
SZ=1000	Biases	-0.003	-0.028	0.018	-0.049	-0.008	(-0.019)	(0.005)	0.029
	Coverage rates	0.959	0.950	0.943	0.936	0.957	0.936	0.939	0.939
Scenario 2		x_1 was complete missing at random (MCAR)							
		$r^*_{x,1i} = \gamma_{x,1,0} + \gamma_{x,1,1}x_{1i} + \zeta_{x,1i}$							
SZ=200	Biases	0.004	-0.072	0.012	-0.091	-0.020	(-0.031)		
	Coverage rates	0.946	0.953	0.955	0.936	0.940	0.909		
SZ=500	Biases	0.000	-0.010	0.001	-0.025	-0.010	(0.035)		
	Coverage rates	0.939	0.949	0.968	0.950	0.963	0.935		
SZ=1000	Biases	0.001	-0.009	0.005	-0.014	-0.006	(-0.031)		
	Coverage rates	0.950	0.959	0.961	0.948	0.949	0.909		
		$r^*_{x,1i} = \gamma_{x,1,0} + \gamma_{x,1,1}x_{1i} + \gamma_{x,1,2}x_{2i} + \gamma_{x,1,3}y_i + \zeta_{x,1i}$							
SZ=200	Biases	0.005	-0.232	0.052	-0.187	-0.012	(0.128)	(0.005)	(-0.008)
	Coverage rates	0.952	0.925	0.958	0.931	0.947	0.863	0.911	0.911
SZ=500	Biases	0.002	-0.125	0.034	-0.089	-0.004	(0.014)	(0.016)	(-0.001)
	Coverage rates	0.944	0.946	0.969	0.939	0.970	0.911	0.922	0.922
SZ=1000	Biases	0.001	-0.087	0.028	-0.055	-0.002	(-0.007)	(0.006)	(-0.001)
	Coverage rates	0.946	0.944	0.946	0.941	0.953	0.916	0.934	0.934
Scenario 3		x_1 was missing due to both x_1 and y (a mixture of MNAR and MAR)							
		$r^*_{x,1i} = \gamma_{x,1,0} + \gamma_{x,1,1}x_{1i} + \zeta_{x,1i}$							
SZ=200	Biases	-0.040	0.328	-0.252	-0.293	-0.086			
	Coverage rates	0.839	0.775	0.886	0.896	0.877			
SZ=500	Biases	-0.042	0.369	-0.268	-0.241	-0.077			
	Coverage rates	0.684	0.480	0.760	0.859	0.817			
SZ=1000	Biases	-0.042	0.373	-0.277	-0.251	-0.076			
	Coverage rates	0.430	0.160	0.520	0.760	0.590			
		$r^*_{x,1i} = \gamma_{x,1,0} + \gamma_{x,1,1}x_{1i} + \gamma_{x,1,2}y_i + \zeta_{x,1i}$							
SZ=200	Biases	-0.007	-0.060	-0.007	-0.150	-0.028			
	Coverage rates	0.949	0.946	0.944	0.931	0.937			
SZ=500	Biases	-0.005	0.008	-0.010	-0.047	-0.014			
	Coverage rates	0.940	0.942	0.957	0.945	0.958			
SZ=1000	Biases	-0.002	-0.002	0.001	-0.019	-0.008			
	Coverage rates	0.952	0.951	0.960	0.937	0.953			
		$r^*_{x,1i} = \gamma_{x,1,0} + \gamma_{x,1,1}x_{1i} + \gamma_{x,1,2}x_{2i} + \gamma_{x,1,3}y_i + \zeta_{x,1i}$							
SZ=200	Biases	-0.010	-0.134	0.042	-0.157	-0.032	0.327	(0.031)	0.120
	Coverage rates	0.959	0.936	0.948	0.942	0.952	0.92	0.935	0.935
SZ=500	Biases	-0.004	-0.080	0.021	-0.098	-0.008	-0.184	(0.034)	0.067
	Coverage rates	0.958	0.940	0.957	0.935	0.945	0.93	0.944	0.944
SZ=1000	Biases	-0.004	-0.029	0.011	-0.046	-0.004	0.067	(0.006)	0.023
	Coverage rates	0.958	0.943	0.946	0.944	0.952	0.941	0.943	0.943

Note: The biases inside the parentheses are absolute biases, and the biases outside the parentheses are relative biases.

Table 2: Real Data Example

Assumption 1						$WS1$, $HS1$, $HS2$ are MAR						
	β_0	β_1	β_2	β_3	β_4	σ^2_ε						
						Full Bayesian						
Estimate	7.438*	0.636*	0.214*	-0.154*	-0.045	9.423						
Standard deviation	2.162	0.055	0.056	0.056	0.092	0.754						
Credible interval	(3.055,	(0.52,	(0.101,	(-0.268,	(-0.209,	(8.223,						
	11.546)	0.737)	0.319)	-0.049)	0.151)	11.183)						
DIC	2198.318											
						Multiple Imputation						
Estimate	7.301*	0.628*	0.210*	-0.157*	-0.026	9.636						
Standard Error	2.196	0.056	0.055	0.058	0.092							
AIC	2206.187			BIC	2230.584							
Assumption 2						$WS1$ and $HS2$ are MAR, and $HS1$ is MNAR						
	β_0	β_1	β_2	β_3	β_4	σ^2_ε	$\gamma_{HS1\|0}$	$\gamma_{HS1\|HS1}$	$\gamma_{HS1\|WS1}$	$\gamma_{HS1\|EDU}$	$\gamma_{HS1\|STR}$	$\gamma_{HS1\|HS2}$
						Full Bayesian						
Estimate	8.132*	0.629*	0.173*	-0.152*	-0.027	9.484	-13.278*	-0.139	0.533*	-0.069	-0.036	-0.084
Standard deviation	2.282	0.056	0.061	0.056	0.092	0.77	2.388	0.099	0.087	0.045	0.077	0.085
Credible interval	(3.543	(0.519	(0.065	(-0.265	(-0.213	(8.276	(-17.795	(-0.319	(0.384	(-0.164	(-0.198	(-0.223
	12.514)	0.74)	0.303)	-0.046)	0.149)	11.309)	-8.389)	0.074)	0.729)	0.014)	0.106)	0.11)
DIC	2201.439											
						Multiple Imputation						
Estimate	8.194*	0.627*	0.18*	-0.156*	-0.033	9.676						
Standard Error	2.257	0.056	0.06	0.056	0.092							
AIC	2208.006			BIC	2232.403							
Assumption 3						$HS1$ and $HS2$ are MAR, and $WS1$ is MNAR						
	β_0	β_1	β_2	β_3	β_4	σ^2_ε	$\gamma_{WS1\|0}$	$\gamma_{WS1\|HS1}$	$\gamma_{WS1\|WS1}$	$\gamma_{WS1\|EDU}$	$\gamma_{WS1\|STR}$	$\gamma_{WS1\|HS2}$
						Full Bayesian						
Estimate	9.118*	0.558*	0.219*	-0.162*	-0.026	9.638	-15.678*	0.616*	-0.1	-0.034	0.06	-0.164*
Standard deviation	2.234	0.057	0.057	0.056	0.093	0.785	2.325	0.092	0.059	0.04	0.076	0.056
Credible interval	(4.944	(0.449	(0.101	(-0.268	(-0.226	(8.404	(-20.183	(0.47	(-0.194	(-0.113	(-0.095	(-0.277
	13.697)	0.672)	0.324)	-0.05)	0.139)	11.478)	-11.083)	0.833)	0.038)	0.043)	0.206)	-0.057)
DIC	2208.460											
						Multiple Imputation						
Estimate	9.643*	0.552*	0.213*	-0.161*	-0.049	9.806						
Standard Error	2.18	0.057	0.058	0.055	0.092							
AIC	2213.795			BIC	2238.191							
Assumption 4						$WS1$ and $HS1$ are MAR, and $HS2$ is MNAR						
	β_0	β_1	β_2	β_3	β_4	σ^2_ε	$\gamma_{HS2\|0}$	$\gamma_{HS2\|HS1}$	$\gamma_{HS2\|WS1}$	$\gamma_{HS2\|EDU}$	$\gamma_{HS2\|STR}$	$\gamma_{HS2\|HS2}$
						Full Bayesian						
Estimate	14.862*	0.548*	0.13	-0.233*	-0.024	14.522	-9.217*	-0.318*	-0.162*	0.032	-0.07	0.655*
Standard deviation	2.606	0.068	0.067	0.063	0.106	1.365	2.321	0.047	0.036	0.043	0.077	0.088
Credible interval	(10.176	(0.417	(-0.007	(-0.364	(-0.246	(12.262	(-13.947	(-0.417	(-0.232	(-0.041	(-0.249	(0.503
	20.381)	0.682)	0.258)	-0.116)	0.174)	17.571)	-4.899)	-0.232)	-0.089)	0.128)	0.051)	0.846)
DIC	2381.857											
						Multiple Imputation						
Estimate	15.385*	0.548*	0.122	-0.238*	-0.033	14.625						
Standard Error	2.584	0.067	0.065	0.063	0.105							
AIC	2385.629			BIC	2410.026							

Note: * indicates $p < 0.05$ in multiple imputation and QBP interval excluding 0 in full Bayesian framework.

Table 3: Real Data Example (continued)

Assumption 5	$HS2$ is MAR, $HS1$ and $WS1$ are MNAR										
	β_0	β_1	β_2	β_3	β_4	σ^2_ε	$\gamma_{HS1\|0}$	$\gamma_{HS1\|HS1}$	$\gamma_{HS1\|WS1}$	$\gamma_{HS1\|EDU}$	$\gamma_{HS1\|STR}$
Full Bayesian											
Estimate	7.786*	0.633*	0.204*	-0.158*	-0.029	9.426	-13.622*	-0.305*	0.655*	-0.081	-0.068
Standard deviation	2.155	0.056	0.054	0.056	0.092	0.749	2.239	0.085	0.081	0.055	0.097
Credible interval	(3.242	(0.519	(0.098	(-0.268	(-0.204	(8.24	(-18.309	(-0.493	(0.511	(-0.187	(-0.276
	11.712)	0.737)	0.309)	-0.05)	0.157)	11.166)	-9.517)	-0.157)	0.83)	0.031)	0.107)
							$\gamma_{WS1\|0}$	$\gamma_{WS1\|HS1}$	$\gamma_{WS1\|WS1}$	$\gamma_{WS1\|EDU}$	$\gamma_{WS1\|STR}$
							-14.592*	-0.301*	0.662*	-0.084	-0.116
							2.37	0.089	0.085	0.057	0.104
							(-19.296	(-0.503	(0.535	(-0.184	(-0.336
							-9.998)	-0.16)	0.872)	0.036)	0.072)
DIC	**2198.275**										
Multiple Imputation											
Estimate	7.181	0.634	0.203	-0.152	-0.017	9.617					
Standard Error	2.209	0.057	0.05	0.057	0.094						
AIC	**2205.353**			BIC	**2229.750**						

Assumption 6	$WS1$ is MAR, $HS1$ and $HS2$ are MNAR											
Full Bayesian												
	β_0	β_1	β_2	β_3	β_4	σ^2_ε	$\gamma_{HS1\|0}$	$\gamma_{HS1\|HS1}$	$\gamma_{HS1,WS1}$	$\gamma_{HS1,EDU}$	$\gamma_{HS1,STR}$	$\gamma_{HS1,HS2}$
Estimate	14.867*	0.563*	0.116	-0.232*	-0.016	15.56	-6.803*	-0.128*	-0.108	0.035	-0.055	0.336*
Standard deviation	2.903	0.081	0.08	0.065	0.11	1.436	1.707	0.059	0.064	0.036	0.066	0.054
Credible interval	(9.209	(0.401	(-0.028	(-0.364	(-0.231	(13.154	(-10.201	(-0.241	(-0.207	(-0.029	(-0.202	(0.248
	20.679)	0.718)	0.285)	-0.109)	0.199)	18.773)	-3.54)	-0.008)	0.044)	0.111)	0.058)	0.461)
							$\gamma_{HS2\|0}$	$\gamma_{HS2\|HS1}$	$\gamma_{HS2\|WS1}$	$\gamma_{HS2\|EDU}$	$\gamma_{HS2\|STR}$	$\gamma_{HS2\|HS2}$
							-9.662*	-0.296*	-0.188*	0.028	-0.112	0.68*
							2.2	0.047	0.04	0.044	0.083	0.088
							(-14.188	(-0.396	(-0.271	(-0.051	(-0.291	(0.527
							-5.586)	-0.21)	-0.114)	0.121)	0.034)	0.872)
DIC	2410.992											
Multiple Imputation												
Estimate	14.855*	0.558*	0.131	-0.236*	-0.019	15.814						
Standard Error	2.866	0.079	0.081	0.064	0.109							
AIC	2419.306			BIC	2443.703							

Assumption 7	$HS1$ is MAR, $WS1$ and $HS2$ are MNAR											
Full Bayesian												
	β_0	β_1	β_2	β_3	β_4	σ^2_ε	$\gamma_{WS1\|0}$	$\gamma_{WS1\|HS1}$	$\gamma_{WS1\|WS1}$	$\gamma_{WS1\|EDU}$	$\gamma_{WS1\|STR}$	
Estimate	10.574*	0.703*	0.088	-0.209*	0.033	13.945	-16.392*	0.541*	-0.139*	-0.012	0.02	
Standard deviation	2.423	0.058	0.066	0.062	0.104	1.277	2.354	0.068	0.053	0.039	0.079	
Credible interval	(5.803	(0.591	(-0.044	(-0.333	(-0.17	(11.883	(-21.025	(0.41	(-0.231	(-0.092	(-0.145	
	15.323)	0.82)	0.216)	-0.09)	0.239)	16.884)	-11.846)	0.676)	-0.021)	0.063)	0.164)	
							$\gamma_{HS2\|0}$	$\gamma_{HS2\|HS1}$	$\gamma_{HS2\|WS1}$	$\gamma_{HS2\|EDU}$	$\gamma_{HS2\|STR}$	$\gamma_{HS2\|HS2}$
							-12.653*	-0.219*	-0.177*	0.022	-0.068	0.68*
							2.206	0.048	0.039	0.042	0.077	0.085
							(-17.062	(-0.32	(-0.258	(-0.056	(-0.234	(0.533
							-8.401)	-0.131)	-0.106)	0.11)	0.07)	0.867)
DIC	2366.208											
Multiple Imputation												
Estimate	10.586*	0.704*	0.086	-0.212*	0.04	14.391						
Standard Error	2.472	0.06	0.07	0.061	0.105							
AIC	2378.813			BIC	2403.210							

Note: * indicates $p < 0.05$ in multiple imputation and QBP interval excluding 0 in full Bayesian framework. The smallest AIC, BIC, and DIC are highlighted in bold.

Figure Captions

Figure 1. Average relative biases of β_3 from the Bayesian latent variable selection model, the misspecified method with an MAR assumption, and the ordinary least squares estimation (OLS) with the original complete data in Simulation Study 1 (MNAR selection process of Y is denoted as $R^2_{r^*_y}$, Y missingness is denoted as P_y, MAR selection process of X_1 is denoted as $R^2_{r^*_{x1}}$, and X_1 missingness is denoted as P_{x1})

Figure 2. Average relative biases of β_0 and β_1 from the Bayesian latent variable selection model, the misspecified method with an MAR assumption, and the ordinary least squares estimation (OLS) with the original complete data in Simulation Study 1 (MNAR selection process of Y is denoted as $R^2_{r^*_y}$, Y missingness is denoted as P_y, MAR selection process of X_1 is denoted as $R^2_{r^*_{x1}}$, and X_1 missingness is denoted as P_{x1})

Figure 3. Average relative biases of β_2 and σ^2_ε from the Bayesian latent variable selection model, the misspecified method with an MAR assumption, and the ordinary least squares estimation (OLS) with the original complete data in Simulation Study 1 (MNAR selection process of Y is denoted as $R^2_{r^*_y}$, Y missingness is denoted as P_y, MAR selection process of X_1 is denoted as $R^2_{r^*_{x1}}$, and X_1 missingness is denoted as P_{x1})

Figure 4. Coverage rates of β_0, β_1, β_2, β_3, and σ^2_ε from the Bayesian latent variable selection model and the misspecified method with an MAR assumption in Simulation Study 1 (MNAR selection process of Y is denoted as $R^2_{r^*_y}$ and Y missingness is denoted as P_y)

Figure 5. Average relative biases of β_3 from the Bayesian latent variable selection model, the misspecified method with an MAR assumption, and the ordinary least squares estimation (OLS) with the original complete data in Simulation Study 2 (MNAR selection process of Y is denoted as $R^2_{r^*_y}$, Y missingness is denoted as P_y, MNAR selection process of X_1 is denoted as $R^2_{r^*_{x1}}$, and X_1 missingness is denoted as P_{x1})

Figure 6. Average relative biases of β_0 and β_1 from the Bayesian latent variable selection model, the misspecified method with an MAR assumption, and the ordinary least squares estimation (OLS) with the original complete data in Simulation Study 2 (MNAR selection process of Y is denoted as $R^2_{r^*_y}$, Y

missingness is denoted as P_y, MNAR selection process of X_1 is denoted as $R^2_{r^*_{x1}}$, and X_1 missingness is denoted as P_{x1})

Figure 7. Average relative biases of β_2 and σ^2_ε from the Bayesian latent variable selection model, the misspecified method with an MAR assumption, and the ordinary least squares estimation (OLS) with the original complete data in Simulation Study 2 (MNAR selection process of Y is denoted as $R^2_{r^*_y}$, Y missingness is denoted as P_y, MNAR selection process of X_1 is denoted as $R^2_{r^*_{x1}}$, and X_1 missingness is denoted as P_{x1})

Figure 8. Coverage rates of β_0, β_1, β_2, β_3, and σ^2_ε from the Bayesian latent variable selection model and the misspecified method with an MAR assumption in Simulation Study 2 (MNAR selection process of Y is denoted as $R^2_{r^*_y}$ and Y missingness is denoted as P_y)

Figure 1: Average relative biases of β_3 from the Bayesian latent variable selection model, the misspecified method with an MAR assumption, and the ordinary least squares estimation (OLS) with the original complete data in Simulation Study 1 (MNAR selection process of Y is denoted as $R^2_{r^*_y}$, Y missingness is denoted as P_y, MAR selection process of X_1 is denoted as $R^2_{r^*_{x1}}$, and X_1 missingness is denoted as P_{x1})

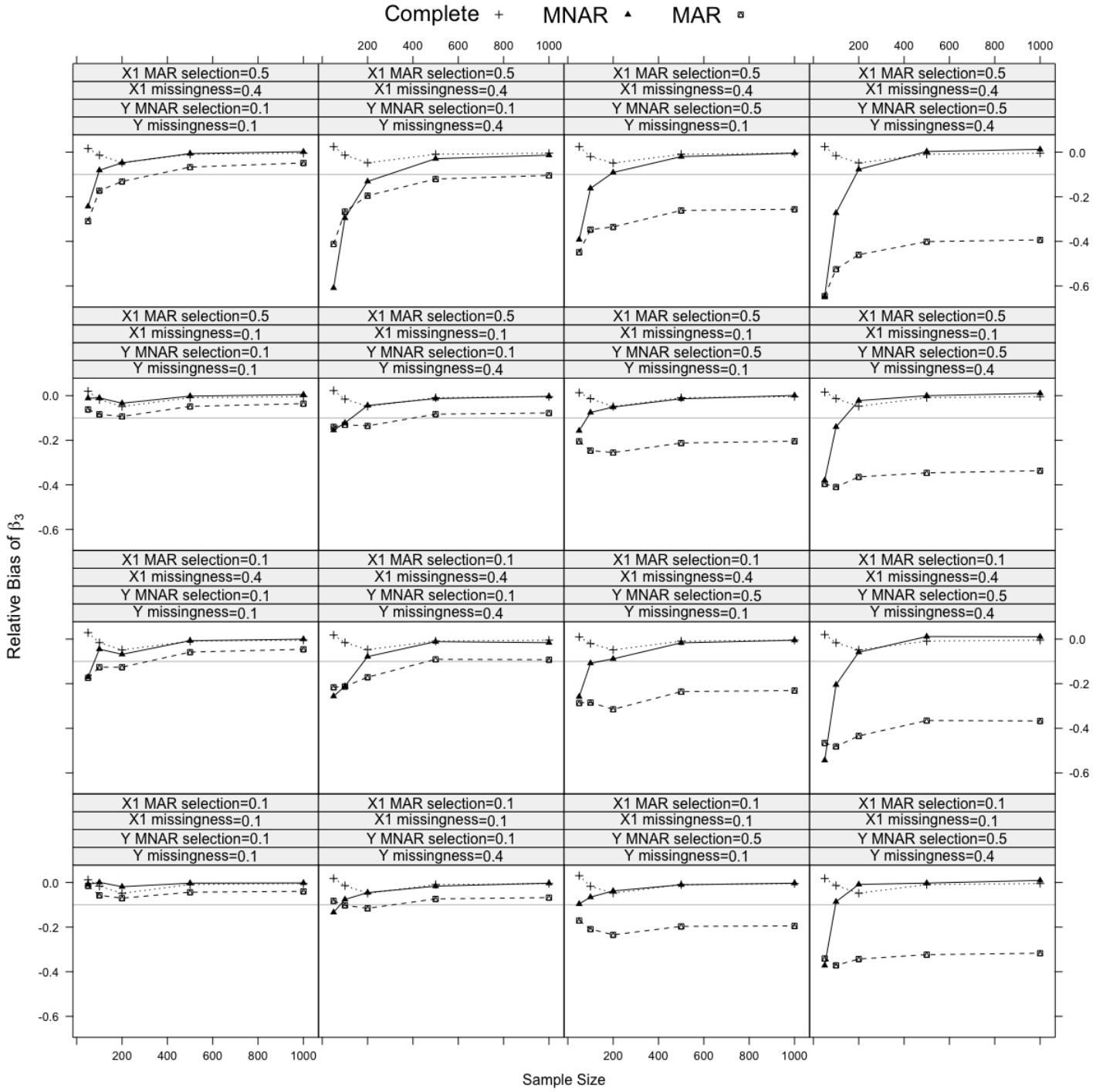

Figure 2: Average relative biases of β_0 and β_1 from the Bayesian latent variable selection model, the misspecified method with an MAR assumption, and the ordinary least squares estimation (OLS) with the original complete data in Simulation Study 1 (MNAR selection process of Y is denoted as $R^2_{r^*_y}$, Y missingness is denoted as P_y, MAR selection process of X_1 is denoted as $R^2_{r^*_{x1}}$, and X_1 missingness is denoted as P_{x1})

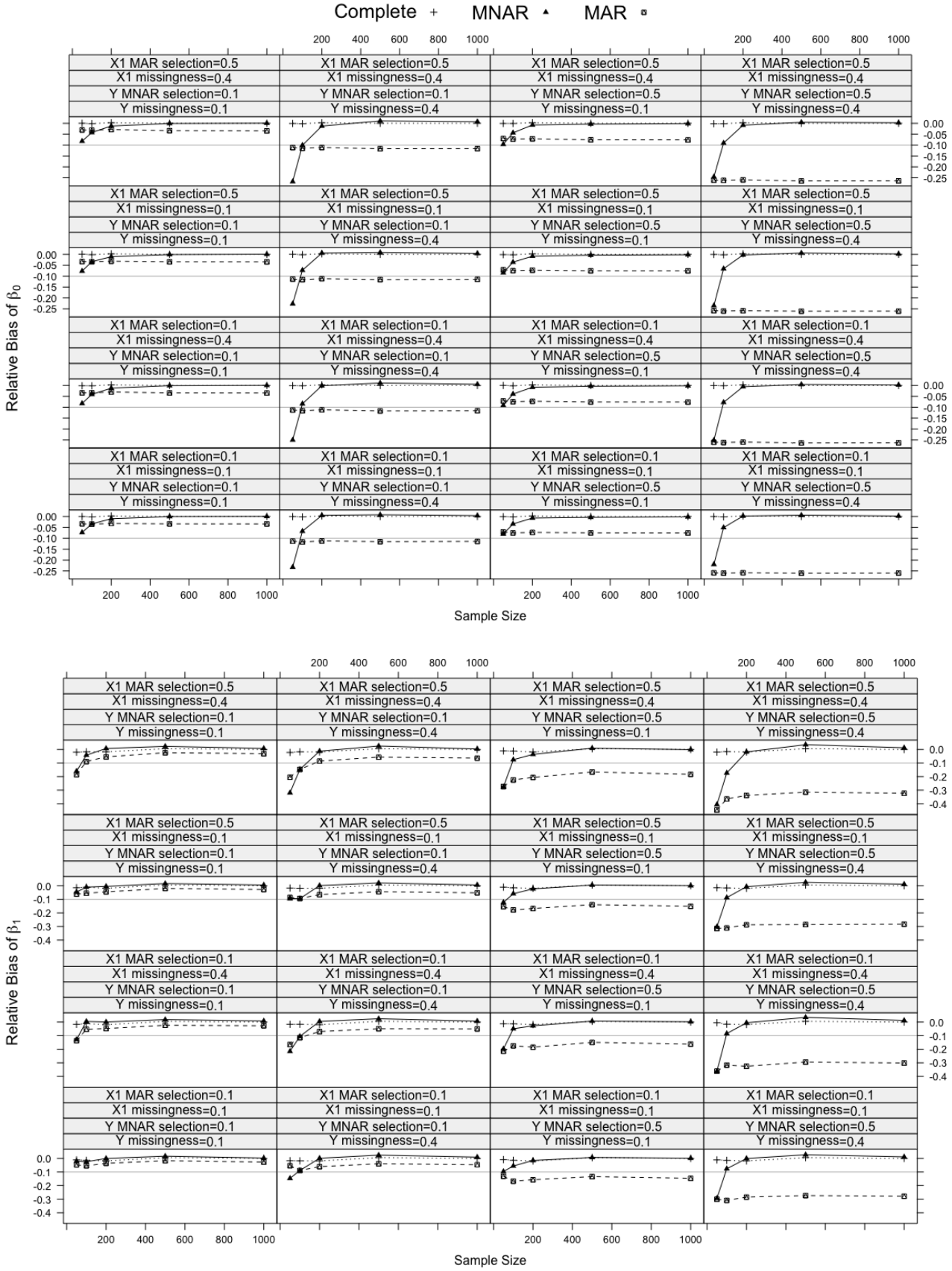

Figure 3: Average relative biases of , β_2 and σ_ε^2 from the Bayesian latent variable selection model, the misspecified method with an MAR assumption, and the ordinary least squares estimation (OLS) with the original complete data in Simulation Study 1 (MNAR selection process of Y is denoted as $R^2_{r_y^*}$, Y missingness is denoted as P_y, MAR selection process of X_1 is denoted as $R^2_{r_{x1}^*}$, and X_1 missingness is denoted as P_{x1})

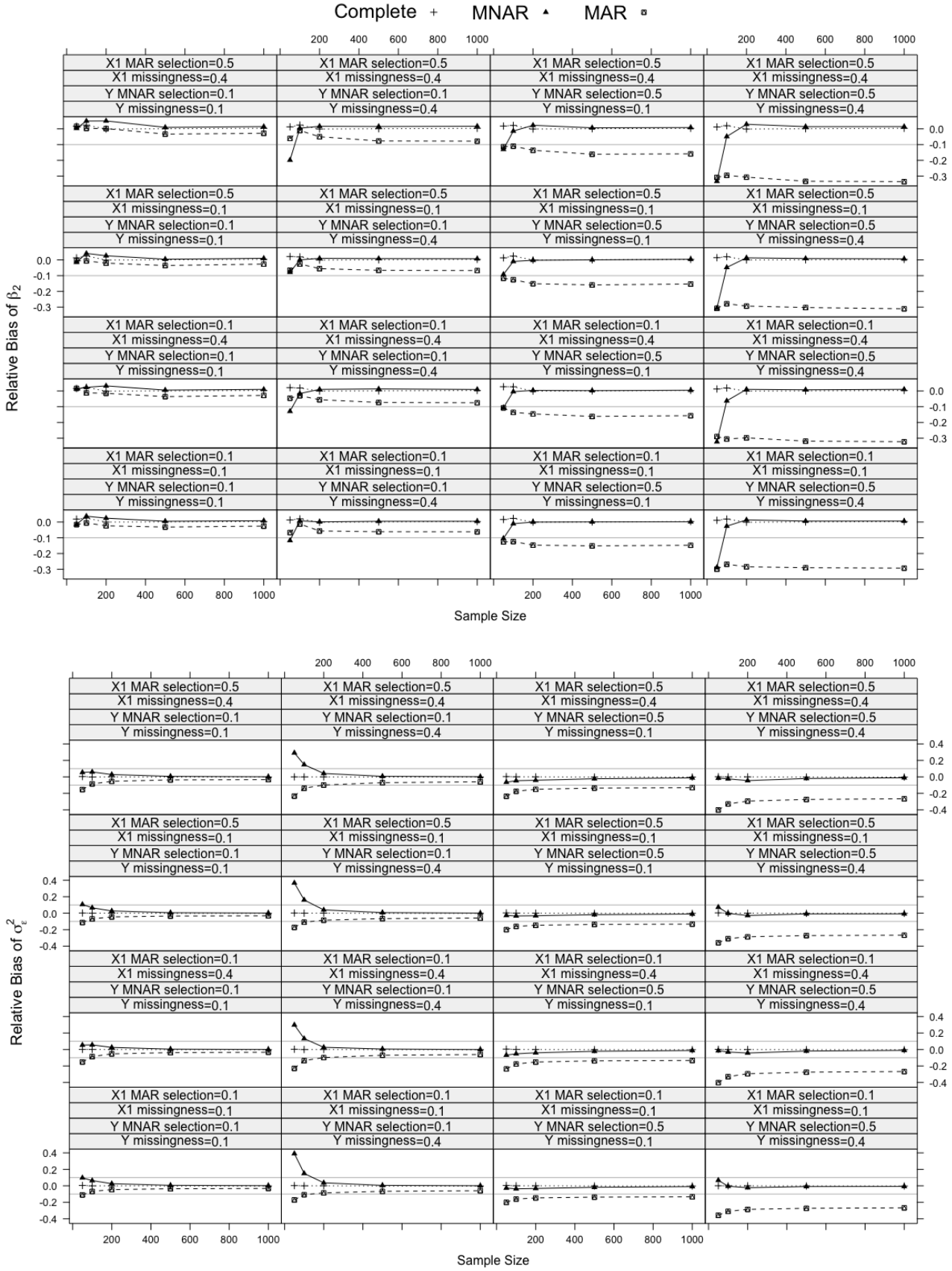

Figure 4: Coverage rates of β_0, β_1, β_2, β_3, and σ_ε^2 from the Bayesian latent variable selection model and the misspecified method with an MAR assumption in Simulation Study 1 (MNAR selection process of Y is denoted as $R^2_{r^*_y}$ and Y missingness is denoted as P_y)

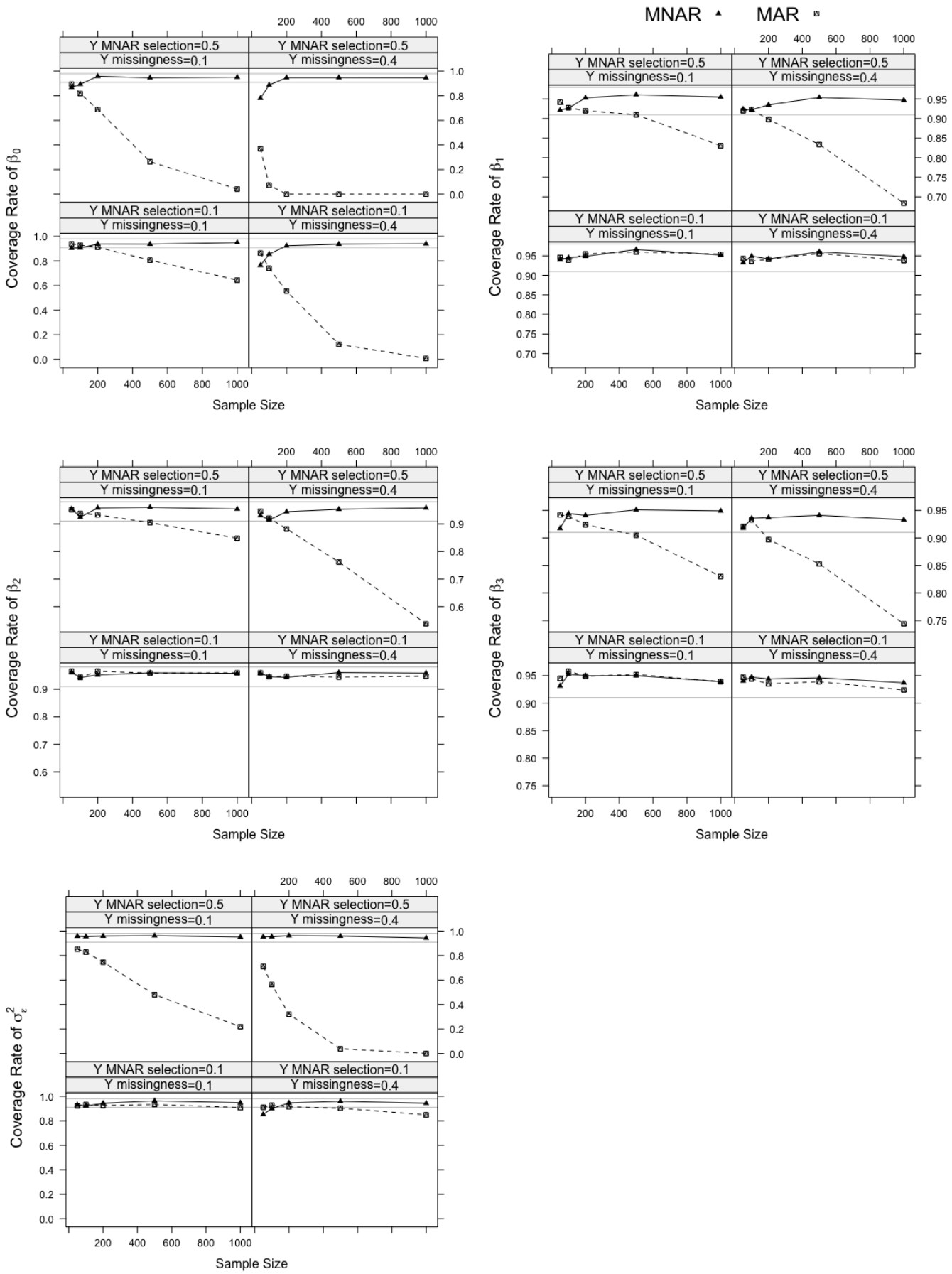

Figure 5: Average relative biases of β_3 from the Bayesian latent variable selection model, the misspecified method with an MAR assumption, and the ordinary least squares estimation (OLS) with the original complete data in Simulation Study 2 (MNAR selection process of Y is denoted as $R^2_{r^*_y}$, Y missingness is denoted as P_y, MNAR selection process of X_1 is denoted as $R^2_{r^*_{x1}}$, and X_1 missingness is denoted as P_{x1})

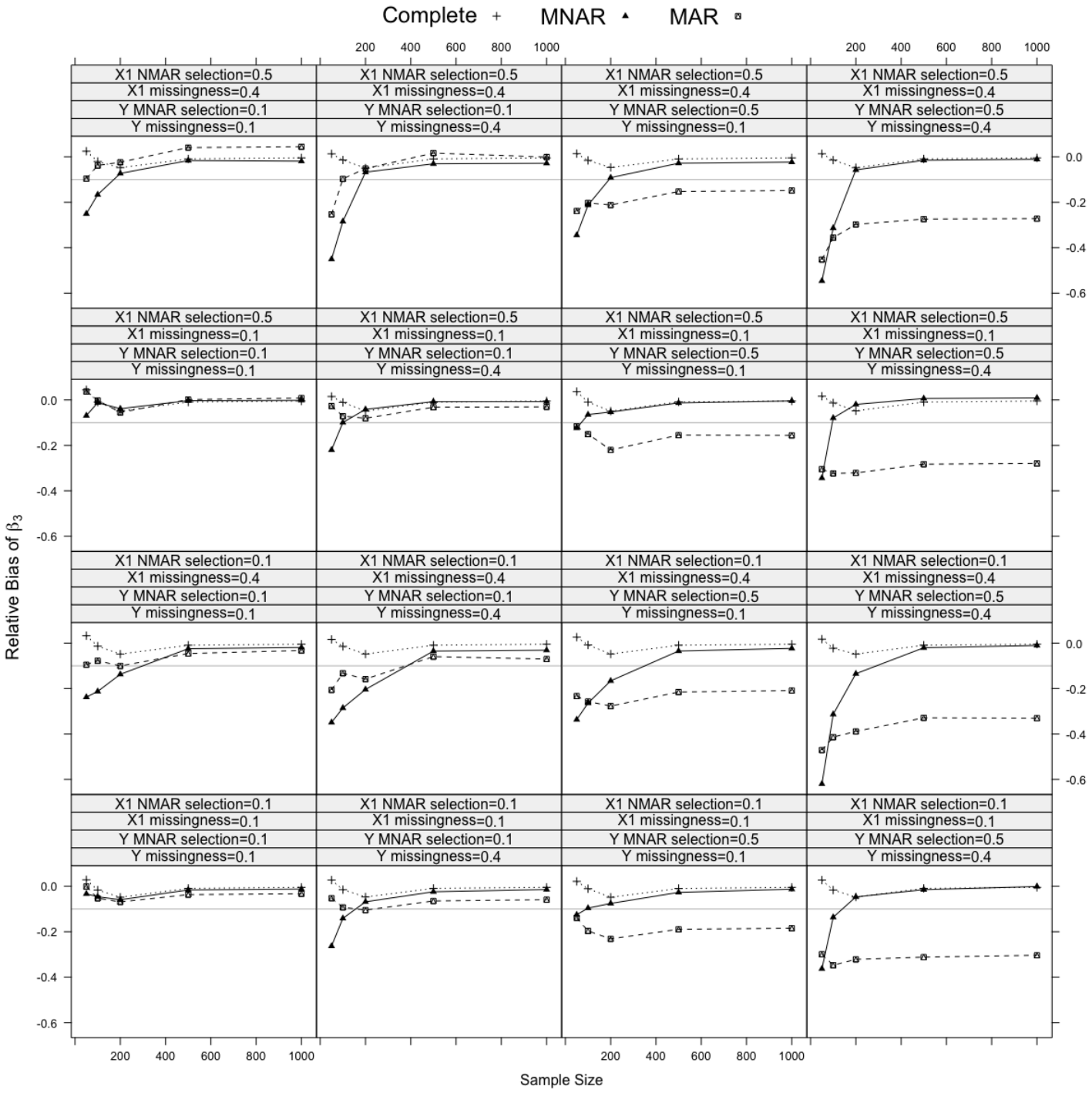

Figure 6: Average relative biases of β_0 and β_1 from the Bayesian latent variable selection model, the misspecified method with an MAR assumption, and the ordinary least squares estimation (OLS) with the original complete data in Simulation Study 2 (MNAR selection process of Y is denoted as $R^2_{r^*_y}$, Y missingness is denoted as P_y, MNAR selection process of X_1 is denoted as $R^2_{r^*_{x1}}$, and X_1 missingness is denoted as P_{x1})

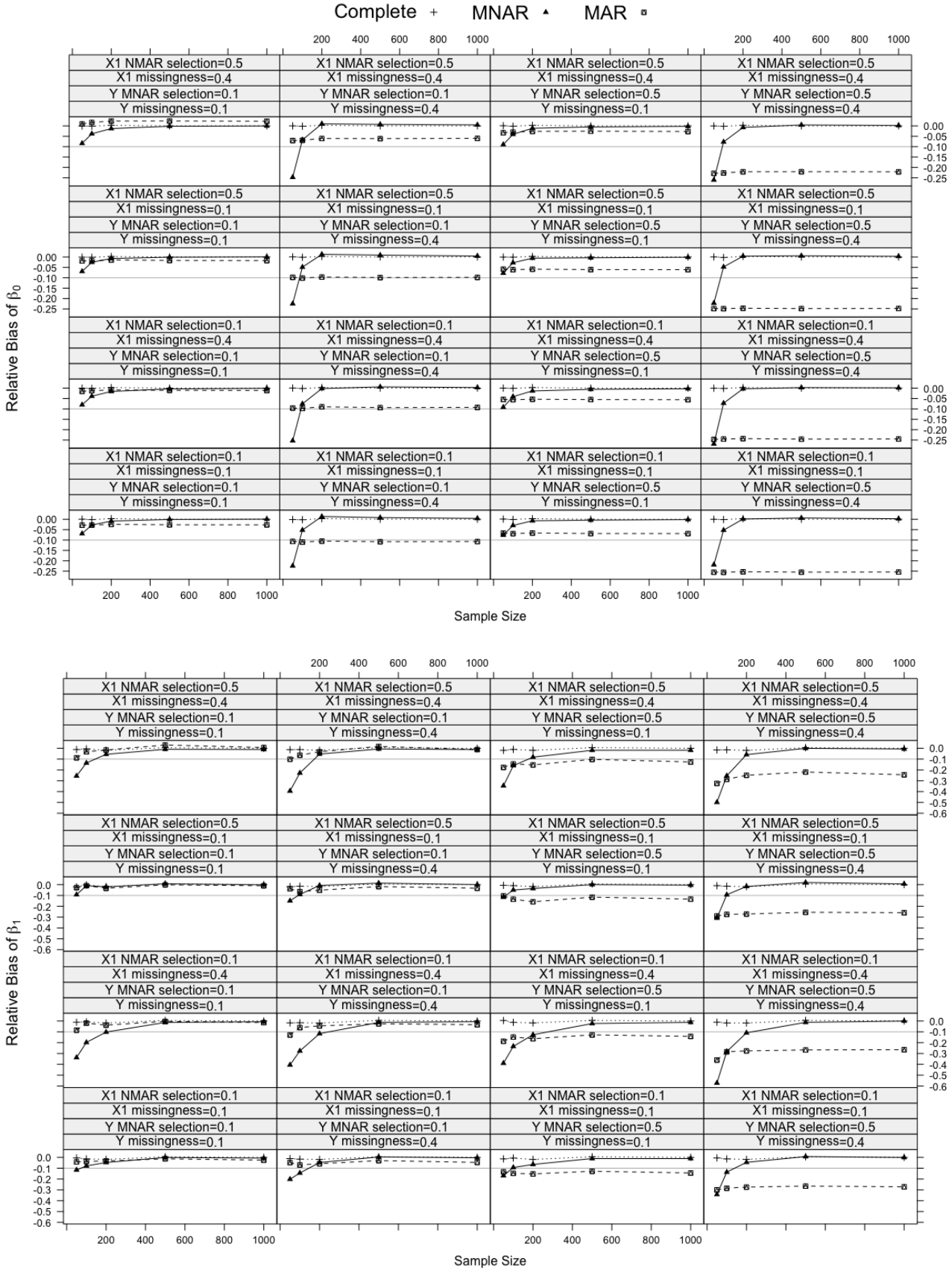

Figure 7: Average relative biases of β_2 and σ_ε^2 from the Bayesian latent variable selection model, the misspecified method with an MAR assumption, and the ordinary least squares estimation (OLS) with the original complete data in Simulation Study 2 (MNAR selection process of Y is denoted as $R^2_{r_y^*}$, Y missingness is denoted as P_y, MNAR selection process of X_1 is denoted as $R^2_{r_{x1}^*}$, and X_1 missingness is denoted as P_{x1})

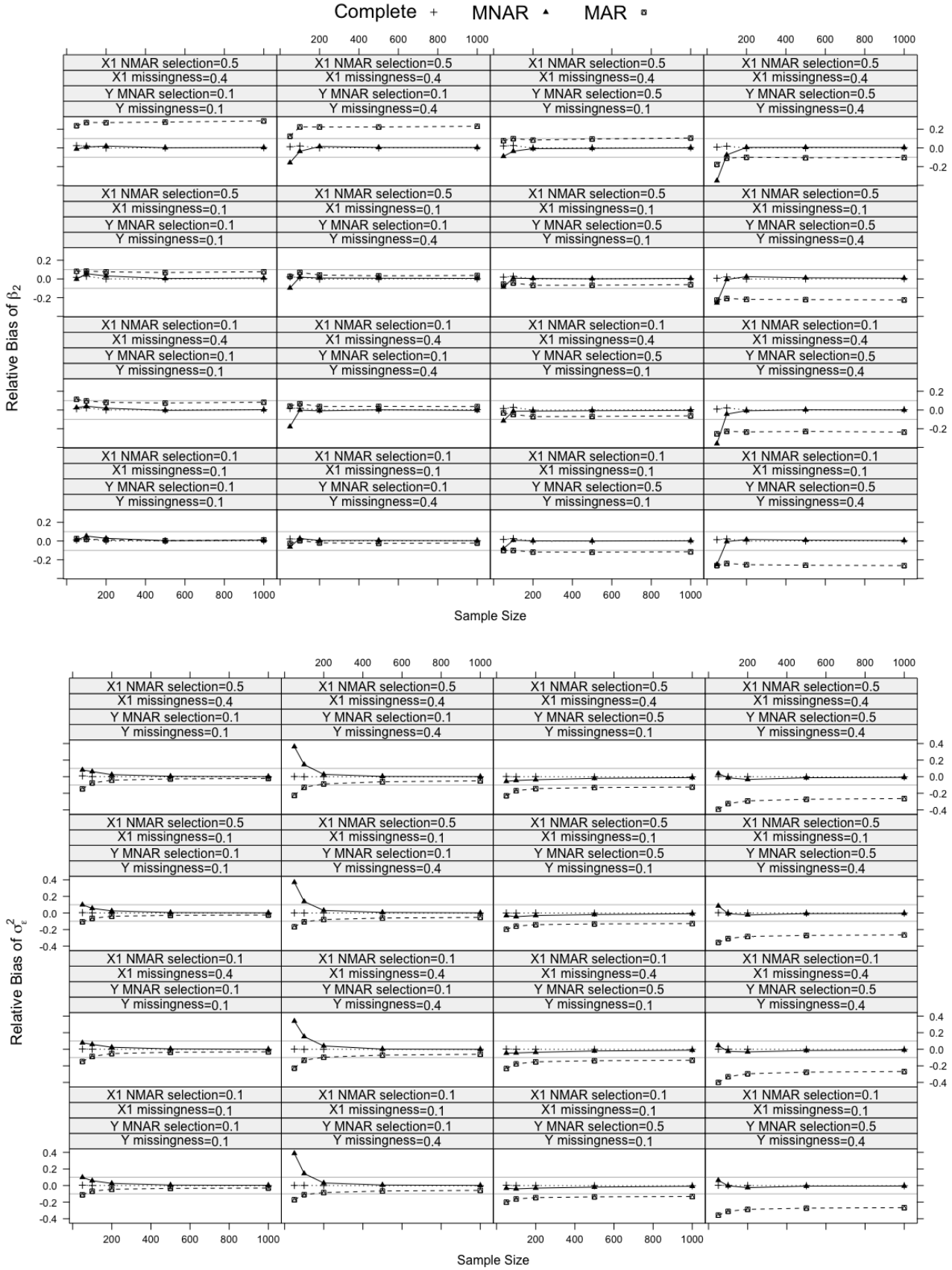

Figure 8: Coverage rates of β_0, β_1, β_2, β_3, and σ_ε^2 from the Bayesian latent variable selection model and the misspecified method with an MAR assumption in Simulation Study 2 (MNAR selection process of Y is denoted as $R^2_{r_y^*}$ and Y missingness is denoted as P_y)