Theory and Inference for Regression Models with Missing Responses and Covariates

Qingxia Chen, **Joseph G. Ibrahim**, **Ming-Hui Chen**, and **Pralay Senchaudhuri**

Qingxia Chen is Assistant Professor, Department of Biostatistics, Vanderbilt University, Nashville, TN 37232, Email: cindy.chen@vanderbilt.edu. Joseph G. Ibrahim is Professor, Department of Biostatistics, University of North Carolina, McGavran-Greenberg Hall, Chapel Hill, NC 27599, Email: ibrahim@bios.unc.edu. Ming-Hui Chen is Professor, Department of Statistics, University of Connecticut, 215 Glenbrook Road, U-4120, Storrs, CT 06269-4120, Email: mhchen@stat.uconn.edu. Pralay Senchaudhuri is Director of Cytel Software Corporation, Cambridge, MA 02139, Email: pralay@cytel.com.

Summary

In this paper, we carry out an in-depth theoretical investigation for inference with missing response and covariate data for general regression models. We assume that the missing data are Missing at Random (MAR) or Missing Completely at Random (MCAR) throughout. Previous theoretical investigations in the literature have focused only on missing covariates or missing responses, but not both. Here, we consider theoretical properties of the estimates under three different estimation settings: complete case analysis (CC), a complete response analysis (CR) that involves an analysis of those subjects with only completely observed responses, and the all case analysis (AC), which is an analysis based on all of the cases. Under each scenario, we derive general expressions for the likelihood and devise estimation schemes based on the EM algorithm. We carry out a theoretical investigation of the three estimation methods in the normal linear model and analytically characterize the loss of information for each method, as well as derive and compare the asymptotic variances for each method assuming the missing data are MAR or MCAR. In addition, a theoretical investigation of bias for the CC method is also carried out. A simulation study and real dataset are given to illustrate the methodology.

1 Introduction

Missing data arise in nearly every type of application in the statistical sciences. Over the past 30 years, there has been an enormous literature on likelihood-based methods of estimation and inference for a wide variety of missing data problems, including missing covariate data in linear models, generalized linear models, generalized linear mixed models, and survival models, as well as missing response data for models of longitudinal data. Since the literature is too vast to list here, we refer the reader to three review articles that discuss various methods for handling missing data: Little (1992), Horton and Laird (1999), and Ibrahim, Chen, Lipsitz, and Herring (2005). There has also been some literature for likelihood-based methods for establishing identifiability and asymptotic properties of estimators in missing covariate problems including Robins and Rotnitzky (1995), Lipsitz, Ibrahim, and Zhao (1999), Herring and Ibrahim (2001), and Chen, Ibrahim, and Shao (2004). There also has been some work on models for

longitudinal data with nonignorable responses, including Baker and Laird (1988), Ibrahim, Chen, and Lipsitz (2001), and Tang, Little, and Raghunathan (2003). There has been some work done on maximum likelihood estimation in the presence of ignorable or nonignorable missing response and/or covariate data in longitudinal models, including Stubbendick and Ibrahim (2003, 2006) and Chen and Ibrahim (2006). However, there has been almost no literature examining theoretical properties of estimators in the presence of *both* MAR responses and covariates in regression models. This type of missing data problem presents many new challenges in estimation and theory that do not arise in missing covariate problems or missing response problems alone.

We refer to a regression problem with missing covariates and responses as a "missing (x, y) problem" throughout. An important issue in a missing (x, y) problem is the contribution to the information matrix of the cases with missing responses alone, the contributions of the cases with missing covariates alone, and the contributions of the cases with missing covariates and responses. In particular, we consider theoretical properties of the estimates under three different estimation settings: complete case analysis (CC), a complete response analysis (CR) that involves an analysis of those subjects with only completely observed responses, and the all case analysis (AC), which is an analysis based on all of the cases. Under each scenario, we derive general expressions for the likelihood and devise estimation schemes based on the EM algorithm. We compare the three estimation methods in the normal linear model and characterize the loss of information for each method as well as derive and compare the asymptotic variances for each method assuming the missing data are MAR. For the linear model, we show that AC analysis has more information than the CR and CC analyses in the sense that the Fisher information for the AC analysis has a greater determinant and trace compared to the Fisher information matrices for the CR and CC analyses, and the CR analysis yields a Fisher information with a greater determinant and trace compared to the Fisher information matrix for the CC analysis. Moreover, we show that the asymptotic variances of the estimates for the CC analyses are larger than the other two methods (CR or AC), and the asymptotic variances for some estimates in the AC analysis are larger than that of the corresponding estimates based on a CR analysis. We also carry out a theoretical investigation of bias for the CC method and analytically show that CC estimates under certain settings are biased.

The rest of this paper is organized as follows. In Section 2, we consider the basic data structure for a regression model with MAR response and/or covariate data. In Section 3, we consider the three analysis methods: CC, CR, and AC. For each method, we give the likelihood function corresponding to the method. Section 4 gets into the heart of the theory and properties of estimators for the three methods and several results are given characterizing the behavior of the Fisher information matrix and asymptotic variances for each method for the normal linear model with missing (x, y). Section 5 examines bias issues for MAR response and covariate data. Section 6 presents a simulation study and real dataset illustrating the theoretical results derived in Section 4. A brief discussion is given in Section 7. In the Appendix A, we devise the computational schemes based on the EM algorithm for obtaining the maximum likelihood estimates (MLE's), and derive E and M-steps of the EM algorithm as well as the observed information matrix based on the observed data using Louis's method for the missing (x, y) problem.

2 Model and Data Structure

2.1 Model

Suppose that $\{(\boldsymbol{x}_i, y_i), i = 1, 2, \ldots, n\}$ are independent observations, where y_i is the response variable, and $x_i = (x_{i1}, \ldots, x_{ip})'$ is a $p \times 1$ random vector of covariates. We specify the joint

distribution of (x_i, y_i) by specifying the conditional distribution of y_i given x_i, denoted $[y_i \mid x_i]$, and the marginal distribution of x_i, denoted $[x_i]$.

We let

$$f(x_i \mid \alpha)$$

denote the joint density for the marginal distribution $[x_i]$, where where α is the vector of model parameters for $i = 1, 2, \ldots, n$. Assume that the distribution function for $[y_i / x_i]$ is of the form

$$f(y_i \mid x_i, \beta, \phi) = f(y_i \mid x_i'\beta, \zeta), \tag{2.1}$$

where $\beta = (\beta_1, \beta_2, \ldots, \beta_p)'$ denotes the $p \times 1$ random vector of regression coefficients, and ζ is the column vector of nuisance parameters. In (2.1), we assume that the distribution $[y_i / x_i]$ depends on x_i and β only through $x'\beta$. If an intercept is included in the model, x_i and β are modified accordingly.

The generalized linear model (GLM) is a special case of (2.1). In the GLM, the conditional density of $[y_i/x_i]$ is given by

$$f(y_i \mid x_i, \beta, \tau) = \exp\{a_i^{-1}\tau(y_i\theta_i - b(\theta_i)) + c(y_i, \tau)\}, \; i = 1, 2, \ldots, n, \tag{2.2}$$

where $\theta_i = \theta(\eta_i)$ is the canonical parameter, $\eta_i = x_i'\beta$, and τ is a dispersion parameter. The functions b and c determine a particular family in the class, such as the binomial, normal, Poisson, etc. The functions $a_i(\tau)$ are commonly of the form $a_i(\tau) = \tau^{-1}w_i^{-1}$, where the w_i's are known weights. Thus, (2.1) reduces to the GLM with

$$f(y_i \mid x_i, \beta, \zeta) = \exp\{a_i\tau(y_i\theta(x_i'\beta) - b(\theta(x_i'\beta))) + c(y_i, \tau)\}$$

and $\zeta = \tau$.

2.2 Missing Data Structures

We consider a general setting in which y_i and some components of x_i may be missing. Let $M_i = \{1 \le l \le p : x_{il} \text{ is missing}\}$, which denotes the set of indices for the ith missing covariates. We also let $\Omega = \{1, 2, \ldots, p\}$ denote the whole index space for x_i. Table 1 gives the general data structure and characterizations of the various missing data patterns in the missing (x, y) problem.

We denote each pattern above by B_j, $j = 1, \ldots, 5$, and refer to B_j as the jth pattern or j^{th} block. B_1 denotes the portion of the data with both y_i and x_i completely observed. In B_2, y_i is observed while each x_i is at least partially missing or completely missing. In B_3, y_i is missing and x_i is completely observed, in B_4, y_i is missing but at least partial x_i is observed, and both y_i and x_i are completely missing in B_5.

Based on the data structure given in Table 1, we use y_i if the ith response is observed and $y_{i,mis}$ if the i_{th} response variable is missing. Also, we write $x_i = (x_{i,\text{obs}}', x_{i,\text{mis}}')'$ where $x_{i,\text{mis}}' = (x_{il}, l \in M_i)$ and $x_{i,obs}$ is the observed portion of x_i.

3 Three Analysis Methods

In this section, we assume that the missing response and the missing covariates are missing at random (MAR). Under the MAR assumption, we only need to model $[y_i, x_i]$. We give the forms of the observed data log-likelihood functions under three analysis methods: complete case (CC) analysis, complete response (CR) analysis, and all case (AC) analysis.

3.1 Complete Case (CC) Analysis

Because standard techniques for regression models require full response and covariate information, one simple way to avoid the problem of missing data is to analyze only those subjects who are completely observed. This method is known as a complete case (CC) analysis.

Based on the data structure displayed in Table 1, the CC analysis uses the portion of data given in Block B_1. Thus, the likelihood function under this method is given by

$$L_{cc}(\theta) = \prod_{i:y_i \text{ observed, } M_i = \theta} f(y_i|x_i,\beta,\phi)f(x_i|\alpha),$$

(3.1)

where $\theta = (\beta, \phi, \alpha)$, and the log-likelihood function is given by

$$l_{cc}(\theta) = \sum_{i:y_i \text{ observed, } M_i = \theta} [\log f(y_i|x_i,\beta,\phi) + \log f(x_i|\alpha)].$$

(3.2)

3.2 Complete Response (CR) Analysis

The complete response cases (CR) analysis is to analyze only those subjects whose responses are completely observed. Thus, in the CR analysis, we only include the portion of data given in Blocks B_1 and B_2 of Table 1. The likelihood function under CR is given by

$$L_{cr}(\theta) = L_{cc} \prod_{i:y_i \text{ observed, } M_i \neq \theta} \left[\int f(y_i|x_i,\beta,\phi)f(x_{i,\text{obs}},x_{i,\text{mis}}|\alpha)dx_{i,\text{mis}}\right],$$

(3.3)

and the log-likelihood function is given by

$$l_{cr}(\theta) = l_{cc} + \sum_{i:y_i \text{ observed, } M_i \neq \theta} \log \left[\int f(y_i|x_i,\beta,\phi)f(x_{i,\text{obs}},x_{i,\text{mis}}|\alpha)dx_{i,\text{mis}}\right].$$

(3.4)

3.3 All Case (AC) Analysis

The all cases (AC) analysis uses the whole data. The likelihood function is given by

$$L_{ac}(\theta) = L_{cr} \left[\prod_{i:y_i \text{ missing, } M_i = \theta} \int f(y_{i,\text{mis}}|x_i,\beta,\phi)f(x_i|\alpha)dy_{i,\text{mis}}\right]$$

$$\times \left[\prod_{i:y_i \text{ missing, } M_i \neq \theta, M_i \neq \Omega} \int\int f(y_{i,\text{mis}}|x_i,\beta,\phi)f(x_{i,\text{obs}}, x_{i,\text{mis}}|\alpha)dy_{i,\text{mis}}dx_{i,\text{mis}}\right]$$

$$\times \left[\prod_{i:y_i \text{ missing, } M_i = \Omega} \int\int f(y_{i,\text{mis}}|x_{i,\text{mis}},\beta,\phi)f(x_{i,\text{mis}}|\alpha)dy_{i,\text{mis}}dx_{i,\text{mis}}\right].$$

(3.5)

Since

$$\int f(y_{i,\mathrm{mis}}|x_i,\beta,\phi)dy_{i,\mathrm{mis}}=1 \text{ and } \int f(x_{i,\mathrm{mis}}|\alpha)dx_{i,\mathrm{mis}}=1,$$

the likelihood function $L_{ac}(\theta)$ reduces to

$$
\begin{aligned}
L_{ac}(\theta) &= L_{cr}\left[\prod_{i:y_i \text{ missing, } M_i=\theta} f(x_i|\alpha)\right] \\
&\times \left[\prod_{i:y_i \text{ missing, } M_i\neq\theta,M_i\neq\Omega} \int f(x_{i,\mathrm{obs}},x_{i,\mathrm{mis}}|\alpha)dx_{i,\mathrm{mis}}\right].
\end{aligned}
\tag{3.6}
$$

Thus, the portion of the data given in Block B_5 of Table 1 does not make any contribution to the likelihood function even under the AC analysis under the MAR assumption.

Using (3.6), the log-likelihood function is given by

$$
\begin{aligned}
l_{ac}(\theta) &= l_{cr}+\left[\sum_{i:y_i \text{ missing, } M_i=\theta} \log f(x_i|\alpha)\right] \\
&+ \left[\sum_{i:y_i \text{ missing},M_i\neq\theta,M_i\neq\Omega} \log \int f(x_{i,\mathrm{obs}},x_{i,\mathrm{mis}}|\alpha)dx_{i,\mathrm{mis}}\right].
\end{aligned}
\tag{3.7}
$$

4 Theoretical Comparisons Between CC, CR, and AC for the Normal Linear Regression Model

In this section, we characterize the properties of the three analysis methods by examining the Fisher information matrix under each method and determining information loss (gain) for each method as well as comparing the asymptotic variances for each method. This comparison allows us to examine the efficiency of each method. To facilitate comparisons in this section, we assume that the missing data are MCAR, since closed-form analytic results for the Fisher information are available in the (x, y) missing problem only under MCAR.

4.1 Simple Linear Regression Model with Missing Responses and Covariates

We first consider a simple normal regression model with a single covariate and unit variances. In this case, we have

$$f(y_i|x_i,\beta,\alpha)=\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(y_i-\beta_0-\beta_1 x_i)^2}{2}\right\} \text{ and } f(x_i|\alpha)=\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(x_i-\alpha)^2}{2}\right\}. \tag{4.1}$$

Write $\theta = (\beta', \alpha)'$. Let $n_j = \#(B_j)$ be the cardinality of B_j for $j = 1, 2, 3$ and $n = n_1 + n_2 + n_3$. For the CC analysis, we have

$$l_{cc}(\theta)=\sum_{i:y_i \text{ observed, } M_i=\theta}\left[-\log(2\pi) - \frac{(y_i-\beta_0-\beta_1 x_i)^2}{2} - \frac{(x_i-\alpha)^2}{2}\right]$$

and the Fisher information matrix is given by

$$I_{cc}(\theta) = -E\left[\frac{\partial^2}{\partial\theta\partial\theta'}l_{cc}(\theta)\right] = E\left[\sum_{i\in B_1}\begin{pmatrix} 1 & x_i & 0 \\ x_i & x_i^2 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right] = n_1\begin{pmatrix} 1 & \alpha & 0 \\ \alpha & 1+\alpha^2 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

(4.2)

For the CR analysis, we have

$$l_{cr}(\theta) = l_{cc}(\theta) + \sum_{i\in B_2}\left\{-\log\sqrt{2\pi} - \frac{1}{2}\log(1+\beta_1^2) - \frac{1}{2(1+\beta_1^2)}(y_i - \beta_0 - \beta_1\alpha)^2\right\}.$$

After some messy algebra, we obtain the Fisher information matrix given by

$$I_{cr}(\theta) = -E\left[\frac{\partial^2}{\partial\theta\partial\theta'}l_{cr}(\theta)\right]$$

$$= n_1\begin{pmatrix} 1 & \alpha & 0 \\ \alpha & 1+\alpha^2 & 0 \\ 0 & 0 & 1 \end{pmatrix} + n_2\begin{pmatrix} \frac{1}{1+\beta_1^2} & \frac{\alpha}{1+\beta_1^2} & \frac{\beta_1}{1+\beta_1^2} \\ \frac{\alpha}{1+\beta_1^2} & \frac{\alpha^2}{1+\beta_1^2}+\frac{2\beta_1^2}{(1+\beta_1^2)^2} & \frac{\alpha\beta_1}{1+\beta_1^2} \\ \frac{\beta_1}{1+\beta_1^2} & \frac{\alpha\beta_1}{1+\beta_1^2} & \frac{\beta_1^2}{1+\beta_1^2} \end{pmatrix}.$$

(4.3)

For the AC analysis, the log-likelihood function is given by

$$l_{ac}(\theta) = l_{cr}(\theta) + \sum_{i:y_i \text{ missing}, M_i=0}\log f(x_i|\alpha) = l_{cr}(\theta) + \sum_{i\in B_3}\left[-\log\sqrt{2\pi} - \frac{1}{2}(x_i - \alpha)^2\right].$$

The corresponding Fisher information matrix is given by

$$I_{ac}(\theta) = -E\left[\frac{\partial^2}{\partial\theta\partial\theta'}l_{ac}(\theta)\right]$$

$$= n_1\begin{pmatrix} 1 & \alpha & 0 \\ \alpha & 1+\alpha^2 & 0 \\ 0 & 0 & 1 \end{pmatrix} + n_2\begin{pmatrix} \frac{1}{1+\beta_1^2} & \frac{\alpha}{1+\beta_1^2} & \frac{\beta_1}{1+\beta_1^2} \\ \frac{\alpha}{1+\beta_1^2} & \frac{\alpha^2}{1+\beta_1^2}+\frac{2\beta_1^2}{(1+\beta_1^2)^2} & \frac{\alpha\beta_1}{1+\beta_1^2} \\ \frac{\beta_1}{1+\beta_1^2} & \frac{\alpha\beta_1}{1+\beta_1^2} & \frac{\beta_1^2}{1+\beta_1^2} \end{pmatrix} + n_3\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

(4.4)

Using (4.2), we have

$$\det(I_{cc}) = n_1^3, \operatorname{tr}(I_{cc}) = n_1(3+\alpha^2), \text{ and } I_{cc}^{-1} = \frac{1}{n_1}\begin{pmatrix} 1+\alpha^2 & -\alpha & 0 \\ -\alpha & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

(4.5)

Observe that

$$I_{cr}(\theta) = \begin{pmatrix} n_1 & n_1\alpha & 0 \\ n_1\alpha & n_1+n_1\alpha^2+\frac{2n_2\beta_1^2}{(1+\beta_1^2)^2} & 0 \\ 0 & 0 & n_1 \end{pmatrix} + \frac{n_2}{1+\beta_1^2}\begin{pmatrix} 1 \\ \alpha \\ \beta_1 \end{pmatrix}\begin{pmatrix} 1 & \alpha & \beta_1 \end{pmatrix} \stackrel{\text{def}}{=} A+uu',$$

where

$$A = \begin{pmatrix} n_1 & n_1\alpha & 0 \\ n_1\alpha & n_1+n_1\alpha^2+\frac{2n_2\beta_1^2}{(1+\beta_1^2)^2} & 0 \\ 0 & 0 & n_1 \end{pmatrix}$$

and $u = \sqrt{\frac{n_2}{1+\beta_1^2}} \begin{pmatrix} 1 & \alpha & \beta_1 \end{pmatrix}'$. After some algebra, we have

$$A^{-1} = \frac{1}{n_1\left[1+\frac{2n_2\beta_1^2}{n_1(1+\beta_1^2)^2}\right]} \begin{pmatrix} 1+\alpha^2+\frac{2n_2\beta_1^2}{n_1(1+\beta_1^2)^2} & -\alpha & 0 \\ -\alpha & 1 & 0 \\ 0 & 0 & 1+\frac{2n_2\beta_1^2}{n_1(1+\beta_1^2)^2} \end{pmatrix}.$$

Note that $u'A^{-1}u = n_2/n_1$. Thus, we have

$$\begin{aligned} \det(I_{cr}) &= \det(A)(1+u'A^{-1}u) = n_1^3\left[1+\frac{2n_2\beta_1^2}{n_1(1+\beta_1^2)^2}\right]\left(1+\frac{n_2}{n_1}\right) \\ &= n_1^3+n_1^2n_2+\frac{2n_1n_2\beta_1^2(n_1+n_2)}{(1+\beta_1^2)^2} = \det(I_{cc})+n_1^2n_2+\frac{2n_1n_2\beta_1^2(n_1+n_2)}{(1+\beta_1^2)^2}, \end{aligned}$$

(4.6)

the trace of I_{cr} is given by

$$\mathrm{tr}(I_{cr}) = \mathrm{tr}(I_{cc})+n_2\left[1+\frac{\alpha^2}{1+\beta_1^2}+\frac{2\beta_1^2}{(1+\beta_1^2)^2}\right],$$

(4.7)

and the inverse matrix of I_{cr} is given by

$$\begin{aligned} I_{cr}^{-1}(\theta) &= A^{-1}-\frac{1}{1+u'A^{-1}u}A^{-1}uu'A^{-1} = A^{-1}-\frac{1}{1+n_2/n_1}A^{-1}uu'A^{-1} \\ &= A^{-1}-\frac{n_2}{n_1(n_1+n_2)(1+\beta_1^2)}\begin{pmatrix} 1 & 0 & \beta_1 \\ 0 & 0 & 0 \\ \beta_1 & 0 & \beta_1^2 \end{pmatrix} \\ &= \frac{1}{n_1}\begin{pmatrix} 1+\frac{\alpha^2}{1+\frac{2n_2\beta_1^2}{n_1(1+\beta_1^2)^2}}-\frac{n_2}{(n_1+n_2)(1+\beta_1^2)} & -\frac{\alpha}{1+\frac{2n_2\beta_1^2}{n_1(1+\beta_1^2)^2}} & -\frac{n_2\beta_1}{(n_1+n_2)(1+\beta_1^2)} \\ -\frac{\alpha}{1+\frac{2n_2\beta_1^2}{n_1(1+\beta_1^2)^2}} & \frac{1}{1+\frac{2n_2\beta_1^2}{n_1(1+\beta_1^2)^2}} & 0 \\ -\frac{n_2\beta_1}{(n_1+n_2)(1+\beta_1^2)} & 0 & 1-\frac{n_2}{n_1+n_2}\frac{\beta_1^2}{1+\beta_1^2} \end{pmatrix}. \end{aligned}$$

(4.8)

Similarly, we can write

$$I_{ac}(\theta) = I_{cr}+\upsilon\upsilon',$$

where $\upsilon = \sqrt{n_3}\begin{pmatrix} 0 & 0 & 1 \end{pmatrix}'$. We have

$$\begin{aligned}
\det(I_{ac}) \quad &= \det(I_{cr})\,(1+\upsilon'I_{cr}^{-1}\upsilon)=\det(I_{cr})+\det(I_{cr})\,\upsilon'I_{cr}^{-1}\upsilon \\
&= \det(I_{cr})+n_3\left[n_1+\frac{2n_2\beta_1^2}{(1+\beta_1^2)^2}\right]\left[n_1+\frac{n_2}{1+\beta_1^2}\right],
\end{aligned}$$

(4.9)

the trace of I_{ac} is given by

$$\mathrm{tr}(I_{ac})=\mathrm{tr}(I_{cr})+n_3,$$

(4.10)

and the inverse matrix of I_{ac} is given by

$$\begin{aligned}
I_{ac}^{-1}(\theta) &= I_{cr}^{-1}-\frac{1}{1+\upsilon'I_{cr}^{-1}\upsilon}I_{cr}^{-1}\upsilon\upsilon'I_{cr}^{-1} \\[6pt]
&= I_{cr}^{-1}-\frac{b*}{n_1}\begin{pmatrix} n_2\beta_1^2 & 0 & -n_2\beta_1[n_1(1+\beta_1^2)+n_2] \\ 0 & 0 & 0 \\ -n_2\beta_1[n_1(1+\beta_1^2)+n_2] & 0 & [n_1(1+\beta_1^2)+n_2]^2 \end{pmatrix} \\[6pt]
&= \frac{1}{n_1}\begin{pmatrix} 1+\dfrac{\frac{\alpha^2}{1+\frac{2n_2\beta_1^2}{n_1(1+\beta_1^2)^2}}-\frac{n_2}{(n_1+n_2)(1+\beta_1^2)}-b*n_2\beta_1^2}{} & -\dfrac{\alpha}{1+\frac{2n_2\beta_1^2}{n_1(1+\beta_1^2)^2}} & -\dfrac{n_2\beta_1}{(n_1+n_2)(1+\beta_1^2)}+b*n_2\beta_1 b_1^* \\[14pt] -\dfrac{\alpha}{1+\frac{2n_2\beta_1^2}{n_1(1+\beta_1^2)^2}} & \dfrac{1}{1+\frac{2n_2\beta_1^2}{n_1(1+\beta_1^2)^2}} & 0 \\[14pt] -\dfrac{n_2\beta_1}{(n_1+n_2)(1+\beta_1^2)}+b*n_2\beta_1 b_1^* & 0 & 1-\dfrac{n_2}{n_1+n_2}\dfrac{\beta_1^2}{1+\beta_1^2}-b*b_1^{*2} \end{pmatrix},
\end{aligned}$$

(4.11)

where $b^*=\dfrac{n_3}{(n_1+n_2)(1+\beta_1^2)[\,n_2n_3+n_1(n_1+n_2+n_3)(1+\beta_1^2)\,]}$ and $b_1^*=n_1(1+\beta_1^2)+n_2$.

Using (4.5) — (4.11), we are led to the following results.

Result 4.1—*Based on either the determinant or the trace of Fisher information matrix, AC yields most gain in information over both CR and CC, and CR gains more information than CC. Specifically, we have*

$$\det(I_{ac})-\det(I_{cr})=n_3\left[n_1+\frac{2n_2\beta_1^2}{(1+\beta_1^2)^2}\right]\left[n_1+\frac{n_2}{1+\beta_1^2}\right]>0,$$
$$\det(I_{cr})-\det(I_{cc})=n_1^2 n_2+\frac{2n_1 n_2\beta_1^2(n_1+n_2)}{(1+\beta_1^2)^2}>0,$$
$$\mathrm{tr}(I_{ac})-\mathrm{tr}(I_{cr})=n_3>0,$$

and

$$\mathrm{tr}(I_{cr})-\mathrm{tr}(I_{cc})=n_2\left[1+\frac{\alpha^2}{1+\beta_1^2}+\frac{2\beta_1^2}{(1+\beta_1^2)^2}\right].$$

The inverse of the Fisher information gives the asymptotic variance and covariance matrix of the MLE's under each analysis method. Now, let $\mathrm{Var}\,(\widehat{\beta}_{j}.)$ and $\mathrm{Var}\,(\hat{\alpha}.)$ denote the asymptotic variances under each of CC, CR, AC. Then, we have the following results.

Result 4.2—*(i) CR leads to smaller asymptotic variances for all parameters than CC. Specifically, we have*

$$
\begin{aligned}
\mathrm{Var}(\widehat{\beta}_{0,cc}) - \mathrm{Var}(\widehat{\beta}_{0,cr}) &= \frac{2n_2\alpha^2\beta_1^2}{n_1[(1+\beta_1^2)^2+2n_2\beta_1^2]} + \frac{n_2}{n_1(n_1+n_2)(1+\beta_1^2)} > 0, \\
\mathrm{Var}(\widehat{\beta}_{1,cc}) - \mathrm{Var}(\widehat{\beta}_{1,cr}) &= \frac{2n_2\beta_1^2}{n_1[n_1(1+\beta_1^2)^2+2n_2\beta_1^2]} > 0, \\
\mathrm{Var}(\widehat{\alpha}_{cc}) - \mathrm{Var}(\widehat{\alpha}_{cr}) &= \frac{n_1 n_2\beta_1^2}{n_1(n_1+n_2)(1+\beta_1^2)} > 0.
\end{aligned}
$$

(ii) AC improves the asymptotic variances for β_0 and α over CR, but not for β_1. Specifically, we have

$$
\begin{aligned}
\mathrm{Var}(\widehat{\beta}_{0,cr}) - \mathrm{Var}(\widehat{\beta}_{0,ac}) &= b^* n_2\beta_1^2 > 0, \\
\mathrm{Var}(\widehat{\beta}_{1,cr}) - \mathrm{Var}(\widehat{\beta}_{1,ac}) &= 0, \\
\mathrm{Var}(\widehat{\alpha}_{cr}) - \mathrm{Var}(\widehat{\alpha}_{ac}) &= b^* b_1^{*2} > 0,
\end{aligned}
$$

where b^* and b_1^* are given in (4.11).

From Result 4.2, the additional information from Block B_3 does improve the standard errors of $\widehat{\beta}_0$ and $\hat{\alpha}$. Surprisingly, the information from Block B_3 does not help improve the standard error of $\widehat{\beta}_1$.

4.2 Multiple Linear Regression Model with Missing Responses and Covariates

To further examine the theoretical relationship among these three analysis methods, we consider a multiple normal linear regression model with $p \geq 2$. For illustrative purposes, it suffices to consider two missing covariates. Specifically, we assume that $x_{i,p-1}$ and $x_{i,p}$ have at least one missing value and x_{i1}, x_{i2}, and $x_{i,p-2}$ are observed in all cases as we have shown in Section 3.3 that the cases in Block 5 do not make any contribution to the log-likelihood function $l_{ac}(\theta)$. For notational convenience, we let $z_{i1} = (1, x_{i1},\ldots,x_{i,p-2})'$. We further assume

$$
\begin{aligned}
y_i|x_i,\beta &\sim N((z_{i1}',x_{i,p-1},x_{ip})'(\beta_0,\beta_1,\beta_2),\sigma^2), \\
x_{i,p-1}|x_i^{(obs)},\alpha_1 &\sim N(z_{i1}'\alpha_1,\tau_1^2),
\end{aligned}
$$

where $\boldsymbol{\alpha}_1 = (\alpha_{10}, \alpha_{11},\ldots,\alpha_{1,p-2})'$, and

$$
x_{ip}|x_i^{(obs)},x_{i,p-1},\alpha_2 \sim N((z_{i1}',x_{i,p-1})(\alpha_{21},\alpha_{22}),\tau_2^2),
$$

where $\boldsymbol{\alpha}_{21} = (\alpha_{20}, \alpha_{21},\ldots, \alpha_{2,p-2})'$. We assume all variances σ^2,τ_1^2, and τ_2^2 are known. For ease of exploration, we choose $\sigma^2=\tau_1^2=\tau_2^2=1$.

In this setting, we need to consider the cases from Blocks 1 to 4.

For Blocks 2 and 4, we assume $B_2=\cup_{k=2}^4 B_{2k}$ and $B_4=\cup_{k=2}^3 B_{4k}$, where

$$B_{j1}=\{i: \text{both } x_{i,p-1} \text{ and } x_{ip} \text{ are observed}\},$$
$$B_{j2}=\{i: x_{i,p-1} \text{ is missing and } x_{ip} \text{ is observed}\},$$
$$B_{j3}=\{i: x_{i,p-1} \text{ is observed and } x_{ip} \text{ is missing}\}$$

for $j = 2, 4$, and $B_{24} = \{i : \text{both } x_{i,p-1} \text{ and } x_{ip} \text{ are missing}\}$. We further let $n_{jk} = \#(B_{jk})$ be the cardinality of B_{jk} for $j = 2, 4$. Then we have $n_2 = \sum_{k=2}^{4} n_{2k}$ and $n_4 = \sum_{k=2}^{3} n_{4k}$.

Define $A_j = \sum_{i \in B_j} z_{i1} z'_{i1}$, where $j = 1, 22, 23, 24, 3, 42, 43$, and $\theta = (\beta'_0, \beta_1, \beta_2, \alpha'_1, \alpha'_{21}, \alpha_{22})'$. For the CC analysis, we have

$$l_{cc}(\theta) = \sum_{i:y_i \text{ observed}, M_i=0} \left[-\tfrac{3}{2}\log(2\pi) - \frac{(y_i - z'_{i1}\beta_0 - \beta_1 x_{i,p-1} - \beta_2 x_{ip})^2}{2} \right.$$
$$\left. - \frac{(x_{i,p-1} - z_{i1}\alpha_1)^2}{2} - \frac{(x_{ip} - z_{i1}\alpha_{21} - x_{i,p-1}\alpha_{22})^2}{2} \right]$$

and the Fisher information matrix is given by

$$I_{cc}(\theta) = \begin{pmatrix} A_1 & A_1\alpha_1 & A_1\tilde{\alpha}_{12} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \alpha'_1 A_1 & n_1 + \alpha'_1 A_1 \alpha_1 & n_1\alpha_{22} + \alpha'_1 A_1 \tilde{\alpha}_{12} & \mathbf{0} & \mathbf{0} & 0 \\ \tilde{\alpha}'_{12} A_1 & n_1\alpha_{22} + \alpha'_1 A_1 \tilde{\alpha}_{12} & n_1(1+\alpha_{22}^2) + \tilde{\alpha}'_{12} A_1 \tilde{\alpha}_{12} & 0 & 0 & 0 \\ \mathbf{0} & 0 & 0 & A_1 & 0 & 0 \\ \mathbf{0} & 0 & 0 & 0 & A_1 & A_1\alpha_1 \\ \mathbf{0} & 0 & 0 & 0 & \alpha'_1 A_1 & n_1 + \alpha'_1 A_1 \alpha_1 \end{pmatrix},$$

where $\tilde{\alpha}_{12} = \alpha_{21} + \alpha_1 \alpha_{22}$.

For the CR analysis, we have

$$l_{cr}(\theta) = l_{cc}(\theta)$$
$$+ \sum_{i:y_i \text{ observed}, M_i \neq 0} \log \left[\int f(y_i|x_i, \beta) f(x_{i,p}|z_{i1}, x_{i,p-1}, \alpha_2) f(x_{i,p-1}|z_{i1}, \alpha_1) \, dx_{i,\text{mis}} \right]$$
$$= l_{cc}(\theta) + l_{cc,1}(\theta) + l_{cc,2}(\theta),$$

where

$$l_{cc,1}(\theta)$$
$$= \sum_{i \in B_{22}} \left\{ -\log(2\pi) - \frac{\log(1+\beta_1^2 + \alpha_{22}^2)}{2} - [(\alpha_{22}(y_i - z'_{i1}\beta_0 - x_{ip}\beta_2) - \beta_1(x_{ip} - z'_{i1}\alpha_{21}))^2 \right.$$
$$\left. + (y_i - z'_{i1}\beta_0 - \beta_1 z'_{i1}\alpha_1 - x_{ip}\beta_2)^2 + (x_{ip} - z'_{i1}\alpha_{21} - \alpha_{22}z'_{i1}\alpha_1)^2]/2(1+\beta_1^2 + \alpha_{22}^2) \right\}$$

and

$$l_{cc,2}(\theta)$$

$$= \sum_{i\in B_{23}} \left\{ -\log(2\pi) - \frac{1}{2}\log(1+\beta_2^2) - \frac{[(y_i - z_{i1}'\beta_0 - x_{i,p-1}\beta_1 - \beta_2(z_{i1}'\alpha_{21}+x_{i,p-1}\alpha_{22}))]^2}{2(1+\beta_2^2)} \right.$$

$$\left. - \frac{(x_{i,p-1}-z_{i1}'\alpha_1)^2}{2} \right\} + \sum_{i\in B_{24}} \left\{ -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log((\beta_1+\beta_2\alpha_{22})^2+1+\beta_2^2) \right.$$

$$\left. - \frac{[(y_i-z_{i1}'\beta_0-\beta_2 z_{i1}'\alpha_1-(\beta_1+\beta_2\alpha_{22})z_{i1}'\alpha_1]^2}{2((\beta_1+\beta_2\alpha_{22})^2+1+\beta_2^2)} \right\}.$$

After some messy algebra, we obtain the Fisher information matrix given by

$$I_{cr}(\theta) = -E\left[\frac{\partial^2}{\partial\theta\partial\theta'} l_{cr}(\theta) \right] = I_{cr}(\theta) + I_{cr,1}(\theta) + n_{22}I_{cr,2}(\theta) + n_{23}I_{cr,3} + n_{24}I_{cr,4},$$

where

$$I_{cr,1}(\theta) = \begin{pmatrix}
M_1 & M_1\alpha_1 & M_1\tilde{\alpha}_{12} & M_2 & M_3 & M_3\alpha_1 \\
\alpha_1'M_1 & \alpha_1'M_1\alpha_1 & \alpha_1'M_1\tilde{\alpha}_{12} & \alpha_1'M_2 & \alpha_1'M_3 & \alpha_1'M_3\alpha_1 \\
\tilde{\alpha}_{12}'M_1 & \tilde{\alpha}_{12}'M_1\alpha_1 & \tilde{\alpha}_{12}'M_1\tilde{\alpha}_{12} & \tilde{\alpha}_{12}'M_2 & \tilde{\alpha}_{12}'M_3 & \tilde{\alpha}_{12}'M_3\alpha_1 \\
M_2 & M_2\alpha_1 & M_2\tilde{\alpha}_{12} & M_6 & M_4 & M_4\alpha_1 \\
M_3 & M_3\alpha_1 & M_3\tilde{\alpha}_{12} & M_4 & M_5 & M_5\alpha_1 \\
\alpha_1'M_3 & \alpha_1'M_3\alpha_1 & \alpha_1'M_3\tilde{\alpha}_{12} & \alpha_1'M_4 & \alpha_1'M_5 & \alpha_1'M_5\alpha_1
\end{pmatrix},$$

$$I_{cr,2}(\theta) = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 \\
0 & \frac{\alpha_{22}^2}{\upsilon_{22}}+\frac{2\beta_1^2}{\upsilon_{22}^2} & \frac{\alpha_{22}(1+\alpha_{22}^2)}{\upsilon_{22}} & 0 & 0 & -\frac{\alpha_{22}\beta_1}{\upsilon_{22}}+\frac{2\alpha_{22}\beta_1}{\upsilon_{22}^2} \\
0 & \frac{\alpha_{22}(1+\alpha_{22}^2)}{\upsilon_{22}} & \frac{(1+\alpha_{22}^2)^2}{\upsilon_{22}} & 0 & 0 & \frac{\beta_1(1-\alpha_{22}^2)}{\upsilon_{22}} \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & -\frac{\alpha_{22}\beta_1}{\upsilon_{22}}+\frac{2\alpha_{22}\beta_1}{\upsilon_{22}^2} & \frac{\beta_1(1-\alpha_{22}^2)}{\upsilon_{22}} & 0 & 0 & \frac{\beta_1^2}{\upsilon_{22}}+\frac{2\alpha_{22}^2}{\upsilon_{22}^2}
\end{pmatrix},$$

$$I_{cr,3}(\theta) = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 \\
0 & \frac{1}{\upsilon_{23}} & \frac{\alpha_{22}}{\upsilon_{23}} & 0 & 0 & \frac{\beta_2}{\upsilon_{23}} \\
0 & \frac{\alpha_{22}}{\upsilon_{23}} & \frac{\alpha_{22}^2}{\upsilon_{23}}+\frac{2\beta_2^2}{\upsilon_{23}^2} & 0 & 0 & \frac{\beta_2\alpha_{22}}{\upsilon_{23}} \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & \frac{\beta_2}{\upsilon_{23}} & \frac{\beta_2\alpha_{22}}{\upsilon_{23}} & 0 & 0 & \frac{\beta_2^2}{\upsilon_{23}}
\end{pmatrix},$$

$$I_{cr,4}(\theta)=\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{2(\beta_1+\beta_2\alpha_{22})^2}{\upsilon_{24}^2} & \frac{2(\beta_1+\beta_2\alpha_{22})\upsilon_{25}}{\upsilon_{24}^2} & 0 & 0 & \frac{2\beta_2(\beta_1+\beta_2\alpha_{22})^2}{\upsilon_{24}^2} \\ 0 & \frac{2(\beta_1+\beta_2\alpha_{22})\upsilon_{25}}{\upsilon_{24}^2} & \frac{2\upsilon_{25}^2}{\upsilon_{24}^2} & 0 & 0 & \frac{2\beta_2(\beta_1+\beta_2\alpha_{22})\upsilon_{25}}{\upsilon_{24}^2} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{2\beta_2(\beta_1+\beta_2\alpha_{22})^2}{\upsilon_{24}^2} & \frac{2\beta_2(\beta_1+\beta_2\alpha_{22})\upsilon_{25}}{\upsilon_{24}^2} & 0 & 0 & \frac{2\beta_2^2(\beta_1+\beta_2\alpha_{22})^2}{\upsilon_{24}^2} \end{pmatrix},$$

$$\upsilon_{22}=1+\beta_1^2+\alpha_{22}^2, \upsilon_{23}=1+\beta_2^2, \upsilon_{24}=(\beta_1+\beta_2\alpha_{22})^2+1+\beta_2^2, \upsilon_{25}=\beta_2+(\beta_1+\beta_2\alpha_{22})\alpha_{22}, M_1=\frac{1+\alpha_{22}^2}{\upsilon_{22}}A_{22}+\frac{1}{\upsilon_{23}}A_{23}+\frac{1}{\upsilon_{24}}A_{24}, M_2=\frac{\beta_1}{\upsilon_{22}}A$$
.

For the AC analysis, the log-likelihood function is given by

$$
\begin{aligned}
&l_{ac}(\theta) \\
={}& l_{cr}(\theta)+\sum_{i\in B_3}\left\{-\log(2\pi)-\frac{(x_{i,p-1}-z_{i1}'\alpha_1)^2}{2}-\frac{(x_{ip}-z_{i1}'\alpha_{21}-x_{i,p-1}\alpha_{22})^2}{2}\right\} \\
&+\sum_{i\in B42}\left\{-\frac{\log(2\pi)}{2}-\frac{1}{2}\log(1+\alpha_{22}^2)-\frac{(x_{ip}-z_{i1}'\alpha_{21}-\alpha_{22}^2 z_{i1}'\alpha_1)^2}{2(1+\alpha_{22}^2)}\right\}+ \\
&+\sum_{i\in B43}\left\{-\frac{1}{2}\log(2\pi)-\frac{1}{2}(x_{i,p-1}-z_{i1}'\alpha_1)^2\right\}
\end{aligned}
$$

The corresponding Fisher information matrix is given by

$$
\begin{aligned}
I_{ac}(\theta)={}&I_{cr}(\theta) \\
&+\begin{pmatrix} \mathbf{0}_{(p+1)\times(p+1)} & \mathbf{0}_{(p+1)\times(p-1)} & \mathbf{0}_{(p+1)\times(p-1)} & 0 \\ \mathbf{0}_{(p-1)\times(p+1)} & A_3+\frac{\alpha_{22}^2}{\upsilon_{42}}A_{42}+A_{43} & \frac{\alpha_{22}}{\upsilon_{42}}A_{42} & \frac{\alpha_{22}}{\upsilon_{42}}A_{42}\alpha_1 \\ \mathbf{0}_{(p-1)\times(p+1)} & \frac{\alpha_{22}}{\upsilon_{42}}A_{42} & A_3+\frac{1}{\upsilon_{42}}A_{42} & A_3\alpha_1+\frac{1}{\upsilon_{42}}A_{42}\alpha_1 \\ \mathbf{0}_{1\times(p+1)} & \frac{\alpha_{22}}{\upsilon_{42}}\alpha_1'A_{42} & \alpha_1'A_3+\frac{1}{\upsilon_{42}}\alpha_1'A_{42} & i_{22} \end{pmatrix},
\end{aligned}
$$

(4.12)

where $\upsilon_{42}=1+\alpha_{22}^2$ and $i_{22}=n_3+\alpha_1'A_3\alpha_1+n_{42}\frac{2\alpha_{22}^2}{\upsilon_{42}^2}+\frac{1}{\upsilon_{42}}\alpha_1'A_{42}\alpha_1$.

For ease of exploration, we choose p=2, in other words, the completely observed covariates only include the intercept.

Result 4.3—*(i) When $n_{23}=n_{24}=0$, CR leads to smaller asymptotic variances for β_1 and β_2 than CC. Specifically, we have*

$$
\begin{aligned}
&\mathrm{Var}(\widehat{\beta}_{1,cc})-\mathrm{Var}(\widehat{\beta}_{1,cr}) \\
={}&\frac{2n_{22}\beta_1^2[\upsilon_{42}(n_1+n_{22})(n_1\upsilon_{42}^2+2n_{22}\alpha_{22}^2)+n_1^2\upsilon_{42}^2\beta_1^2+n_1n_{22}(1+\alpha_{22}^4)\beta_1^2]}{d_{11}}>0, \\
&\mathrm{Var}(\widehat{\beta}_{2,cc})-\mathrm{Var}(\widehat{\beta}_{2,cr}) \\
={}&\frac{4n_{22}^3\alpha_{22}^2\beta_1^2+n_1^2n_{22}\upsilon_{22}^2+2n_1n_{22}\upsilon_{22}\alpha_{22}^2\beta_1^2(n_1+n_{22})+2n_1n_{22}^2(\alpha_{22}^2+\beta_1^2+3\alpha_{22}^2\beta_1^2)}{d_{11}}>0,
\end{aligned}
$$

where $d_{11}=n_1^2(n_1+n_{22})\upsilon_{22}[n_1\upsilon_{22}^2+2n_{22}(\beta_1^2+\alpha_{22}^2+\beta_1^2\alpha_{22}^2)]+2n_1n_{22}^2\beta_1^2\alpha_{22}^2(3n_1+2n_{22}).$

(ii) When $n_{23} = n_{24} = 0$, AC leads to smaller asymptotic variances for β_1 and β_2 than CR. Specifically, we have

$$\mathrm{Var}(\widehat{\beta}_{1,cr}) - \mathrm{Var}(\widehat{\beta}_{1,ac})=\frac{4n_1n_{22}^2\alpha_{22}^2\beta_1^6(n_3\upsilon_{42}^2+2n_{42}\alpha_{22}^2)\,[\upsilon_{42}(n_1+n_{22})+n_1\beta_1^2]^2}{d_{13}} >0,$$

$$\mathrm{Var}(\widehat{\beta}_{2,cr}) - \mathrm{Var}(\widehat{\beta}_{2,ac})=\frac{n_1n_{22}^2\beta_1^2(n_3\upsilon_{42}^2+2n_{42}\alpha_{22}^2)\,[2n_{22}\beta_1^2-n_1\alpha_{22}^4+n_1(1+\beta_1^2)^2]^2}{d_{13}} >0,$$

where

$d_{13}=d_{11}\{2n_3n_{22}^2\upsilon_{42}^4\beta_1^2+2n_{22}\upsilon_{42}^2\alpha_{22}^2(n_{22}+n_{42})(2n_{22}\beta_1^2+n_1\upsilon_{42})+n_1^3\upsilon_{42}^2\upsilon_{22}^3+n_1n_{22}\upsilon_{42}[n_3\upsilon_{42}^4+2\upsilon_{42}\beta_1^2(2n_3\upsilon_{42}^2+n_{22}+\alpha_{22}^2(n_{22}\upsilon$
.

Result 4.4—(i) *When $n_{22} = n_{24} = 0$, CR leads to smaller asymptotic variances for β_1 and β_2 than CC. Specifically, we have*

$$\mathrm{Var}(\widehat{\beta}_{1,cc}) - \mathrm{Var}(\widehat{\beta}_{1,cr})=\frac{n_1n_{23}\upsilon_{23}^2+2n_{23}\beta_2^2[n_{23}+(n_1+n_{23})\upsilon_{23}\alpha_{22}^2]}{n_1(n_1+n_{23})\upsilon_{23}(2n_{23}\beta_2^2+n_1\upsilon_{23}^2)} >0,$$

$$\mathrm{Var}(\widehat{\beta}_{2,cc}) - \mathrm{Var}(\widehat{\beta}_{2,cr})=\frac{2n_{23}\beta_2^2}{n_1(2n_{23}\beta_2^2+n_1\upsilon_{23}^2)} >0.$$

(ii) When $n_{22} = n_{24} = 0$, AC improves the asymptotic variances for $\boldsymbol{\beta}_1$ over CR, but not for $\boldsymbol{\beta}_1$. Specifically, we have

$$\mathrm{Var}(\widehat{\beta}_{1,cr}) - \mathrm{Var}(\widehat{\beta}_{1,ac})=\frac{n_{23}^2(\upsilon_{42}^2n_3+2\alpha_{22}^2n_{42})\beta_2^2}{d_{23}} >0,$$

$$\mathrm{Var}(\widehat{\beta}_{2,cr}) - \mathrm{Var}(\widehat{\beta}_{2,ac})=0,$$

where $d_{23}=n_1(n_1+n_{23})\upsilon_{23}[n_1(n_1+n_{23})\upsilon_{23}\upsilon_{42}^2+(2n_{42}\alpha_{22}^2+n_3\upsilon_{42}^2)(n_1\upsilon_{23}+n_{23})].$

Result 4.5—(i) *When $n_{22} = n_{23} = 0$, CR leads to smaller asymptotic variances for $\boldsymbol{\beta}$ than CC. Specifically, we have*

$$\mathrm{Var}(\widehat{\beta}_{1,cc}) - \mathrm{Var}(\widehat{\beta}_{1,cr})=\frac{2n_{24}\beta_1^2}{d_{31}} >0,$$

$$\mathrm{Var}(\widehat{\beta}_{2,cc}) - \mathrm{Var}(\widehat{\beta}_{2,cr})=\frac{2n_{24}\beta_2^2}{d_{31}} >0,$$

where $d_{31}=2n_1n_{24}[\beta_1^2\upsilon_{23}+2\upsilon_{23}\beta_1\beta_2\alpha_{22}+\beta_2^2+\beta_2^2\alpha_{22}^2\upsilon_{23}]+n_1^2\upsilon_{24}^2.$

(ii) When $n_{22} = n_{23} = 0$, AC leads to smaller asymptotic variances for $\boldsymbol{\beta}$ than CR. Specifically, we have

$$\mathrm{Var}(\widehat{\beta}_{1,cr}) - \mathrm{Var}(\widehat{\beta}_{1,ac})=\frac{4n_{24}^2\beta_1^2\beta_2^2(\beta_1+\alpha_{22}\beta_2)^2(2\alpha_{22}^2n_{42}+n_3\upsilon_{42}^2)}{d_{33}} >0,$$

$$\mathrm{Var}(\widehat{\beta}_{2,cr}) - \mathrm{Var}(\widehat{\beta}_{2,ac})=\frac{4n_{24}^2\beta_2^4(\beta_1+\alpha_{22}\beta_2)^2(2\alpha_{22}^2n_{42}+n_3\upsilon_{42}^2)}{d_{33}} >0,$$

where

$$d_{33}=d_{31}[2n_{24}(\beta_1+\beta_2\alpha_{22})^2(n_3v_{42}^2+n_1v_{23}v_{42}^2+2n_{42}\alpha_{22}^2)+v_{42}^2(n_1+n_3)(2n_{24}\beta_2^2+n_1v_{24}^2)+2n_{42}\alpha_{22}^2(2n_{24}\beta_2^2+n_1v_{24}^2)]$$

.

Remark 1—The information in Block B_{43} does not improve the asymptotic variances of the estimates of β_1 and β_2 in all of the three situations considered here.

Remark 2—When $n_{23} = n_{24} = 0$ (Result 4.3), the differences of the asymptotic variances for β_1 and β_2 do not depend on β_2 comparing CR to CC and AC to CR.

Remark 3—When $n_{22} = n_{24} = 0$ (Result 4.4), the differences of the asymptotic variances for β_1 and β_2 do not depend on β_1 comparing CR to CC and AC to CR.

Remark 4—When $n_{22} = n_{23} = 0$ (Result 4.5), the ratios of the asymptotic variances improvement of β_1 versus β_2 equal to β_1^2/β_2^2 for CR versus CC and AC versus CR, i.e.

$$\frac{\mathrm{Var}(\widehat{\beta}_{1,cc}) - \mathrm{Var}(\widehat{\beta}_{1,cr})}{\mathrm{Var}(\widehat{\beta}_{2,cc}) - \mathrm{Var}(\widehat{\beta}_{2,cr})} = \frac{\mathrm{Var}(\widehat{\beta}_{1,cr}) - \mathrm{Var}(\widehat{\beta}_{1,ac})}{\mathrm{Var}(\widehat{\beta}_{2,cr}) - \mathrm{Var}(\widehat{\beta}_{2,ac})} = \frac{\beta_1^2}{\beta_2^2}$$

Remark 5—When $n_{23} = n_{24} = 0$, the differences of the asymptotic variances of β_1 are monotone decreasing function of n_1 for CR versus CC and AC versus CR. Other monotonic properties are listed in Table 2.

5 Analysis of Bias

In this section, we examine bias in the CC situation when the missing data are MAR. For inference with only missing response data, Little and Rubin (2002, page 43.) note without proof that when the data are MAR, the CC estimates are not biased when the missing data mechanism depends only on the covariates and not the response. The estimates, however, are biased if the missing data mechanism depends on the response. We now examine this bias issue in the missing (x, y) problem, where missingness is MAR.

Based on the data structure given in Table 1, define missing data indicators $r_i=(r_{iy},r_{ix}')\prime$ as

$$r_{iy}=\begin{cases} 1 & \text{if } y_i \text{ is observed} \\ 0 & \text{if } y_i \text{ is missing} \end{cases}$$

and

$$r_{ix,j}=\begin{cases} 1 & \text{if the } j^{th} \text{ component of } x_i \text{ is observed,} \\ 0 & \text{if the } j^{th} \text{ component of } x_i \text{ is missing,} \end{cases}$$

for $j = 1, 2,…,p$.

Let $f(r_i|\phi, y_i, x_i)$ denote the distribution of r_i, which may possibly depend on y_i and x_i, where ϕ is the vector of parameters in the r_i model. Under the MAR assumption, following models for r_i are possible:

$$f(r_i|\phi,y_i,x_i)=f(r_i|\phi),\tag{5.1}$$

$$f(r_i|\phi,y_i,x_i)=f(r_{ix}|\phi_1)f(r_{iy}|\phi_2,r_{ix}\circ x_i),\tag{5.2}$$

where $\circ$ denotes the direct product, or

$$f(r_i|\phi,y_i,x_i)=f(r_{iy}|\phi_1)f(r_{ix}|\phi_2,r_{iy}y_i).\tag{5.3}$$

Note that the model specified by (5.1) defines MCAR and some other versions of the MAR models can be considered as well.

Because standard techniques for regression models require full response and covariate information, one simple way to avoid the problem of missing data is to analyze only those subjects who are completely observed. This method is known as a complete case (CC) analysis.

Based on the data structure displayed in Table 1, the CC analysis uses the portion of data given in Block B_1. Thus, the likelihood function under this method is given by

$$L_{cc}(\theta)=\prod_{i:\text{ both } y_i \text{ and } x_i \text{ observed}} f(y_i|x_i,\beta)f(x_i|\alpha)f(r_i|\phi,y_i,x_i),\tag{5.4}$$

where $\theta = (\beta,\alpha,\phi)$, and the log-likelihood function is given by

$$l_{cc}(\theta)=\sum_{i:\text{ both } y_i \text{ and } x_i \text{ observed}} [\log f(y_i|x_i,\beta)+\log f(x_i|\alpha)+\log f(r_i|\phi,y_i,x_i)].\tag{5.5}$$

Under CC, we will make conditional inference given $r_i = \mathbf{1}$, where $\mathbf{1} = (1, 1,\ldots,1)'$. More specifically, we need to consider conditional distribution $[y_i, x_i|r_i = \mathbf{1}]$ in examining biasness of the MLE's and in deriving the Fisher information matrix. We assume throughout that ϕ is distinct from β and α.

Under the model given by (5.1), we have

$$f(y_i,x_i|\beta,\alpha,\phi,r_i=\mathbf{1})=\frac{f(y_i|x_i,\beta)f(x_i|\alpha)f(r_i|\phi)}{\int f(y_i|x_i,\beta)f(x_i|\alpha)f(r_i|\phi)dx_idy_i}=f(y_i|x_i,\beta)f(x_i|\alpha).$$

Thus, under MCAR, the conditional distribution of (y_i, x_i) given $r_i = \mathbf{1}$ is the same as the unconditional distribution and hence, the MLE's of β and α are unbiased or asymptotically consistent under certain usual regularity conditions.

Under MAR with the model given by (5.2) for r_i, we have

$$\begin{aligned}
f(y_i,x_i|\beta,\alpha,\phi,r_i=\mathbf{1}) &= \frac{f(y_i|x_i,\beta)f(x_i|\alpha)f(r_{ix}=\mathbf{1}|\phi_1)f(r_{iy}=\mathbf{1}|\phi_2,x_i)}{\int f(y_i|x_i,\beta)f(x_i|\alpha)f(r_{ix}=\mathbf{1}|\phi_1)f(r_{iy}=\mathbf{1}|\phi_2,x_i)dy_idx_i} \\
&= f(y_i|x_i,\beta) \times \frac{f(x_i|\alpha)f(r_{iy}=\mathbf{1}|\phi_2,x_i)}{\int f(x_i|\alpha)f(r_{iy}=\mathbf{1}|\phi_2,x_i)dx_i}.
\end{aligned}\tag{5.6}$$

From (5.6), it is easy to see that the MLE of $\boldsymbol{\beta}$ is unbiased or asymptotically consistent, but the MLE of $\boldsymbol{\alpha}$ may not be in this case.

Under the MAR with the model given by (5.3) for r_i, we obtain

$$
\begin{aligned}
f(y_i,x_i|\beta,\alpha,\phi,r_i{=}\mathbf{1}) &= \frac{f(y_i|x_i,\beta)f(x_i|\alpha)f(r_{iy}{=}1|\phi_1)f(r_{ix}{=}\mathbf{1}|\phi_2,y_i)}{\int f(y_i|x_i,\beta)f(x_i|\alpha)f(r_{iy}{=}1|\phi_1)f(r_{ix}{=}\mathbf{1}|\phi_2,y_i)dy_idx_i} \\
&= \frac{f(y_i|x_i,\beta)f(x_i|\alpha)f(r_{ix}{=}\mathbf{1}|\phi_2,y_i)}{\int f(y_i|x_i,\beta)f(x_i|\alpha)f(r_{ix}{=}\mathbf{1}|\phi_2,y_i)dy_idx_i}.
\end{aligned}
\tag{5.7}
$$

In this case, the MLE's for both $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are likely to be biased.

To obtain the closed form analytical results for (5.6) and (5.7), we consider the simple normal regression model given by (4.1). For notational simplicity, we assume that both y_i and x_i are observed for $i = 1, 2,\ldots,m$. Let $y_{obs} = (y_1, y_2,\ldots, y_m)'$, $X'_{obs} = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ x_1 & x_2 & \ldots & x_m \end{pmatrix}$, and $r_{obs} = (1, 1,\ldots,1)'$. Then, the MLE of $\boldsymbol{\beta}$ is given by $\widehat{\beta}_{cc}=(X'_{obs}X_{obs})^{-1}X'_{obs}y_{obs}$ and the MLE of α is $\widehat{\alpha}_{cc}=\frac{1}{m}\sum_{i=1}^{m}x_i.$

In (5.6), we assume $f(r_{iy}{=}1|\phi_2,x_i)=\dfrac{\exp(\phi_{20}+\phi_{21}x_i)}{1+\exp(\phi_{20}+\phi_{21}x_i)}.$ Then, (5.6) implies

$$
f(y_i|\beta,\alpha,\phi,x_i,r_i{=}\mathbf{1})=f(y_i|x_i,\beta)=\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(y_i-\beta_0-\beta_1x_i)^2}{2}\right\}
$$

and

$$
f(x_i|\alpha,\phi,r_i{=}\mathbf{1})=\frac{\exp\{-(x_i-\alpha)^2/2\}\left[\frac{\exp(\phi_{20}+\phi_{21}x_i)}{1+\exp(\phi_{20}+\phi_{21}x_i)}\right]}{\int\exp\{-(x_i-\alpha)^2/2\}\left[\frac{\exp(\phi_{20}+\phi_{21}x_i)}{1+\exp(\phi_{20}+\phi_{21}x_i)}\right]dx_i}.
\tag{5.8}
$$

Thus, we have

$$
E[\widehat{\beta}_{cc}|r_{obs}]=E\{E[(X'_{obs}X_{obs})^{-1}X'_{obs}y_{obs}|X_{obs}]|r_{obs}\}=E\{(X'_{obs}X_{obs})^{-1}X'_{obs}X_{obs}\beta|r_{obs}\}=\beta,
$$

which is unbiased. However,

$$
E[\widehat{\alpha}_{cc}|r_{obs}] \neq \alpha,
$$

which may be biased. Also, an analytical derivation of the Fisher information matrix is not possible as the conditional distribution of x_i given $r_i = \mathbf{1}$ involves an analytically intractable integral.

Under MAR given by (5.3), we assume a logistic regression model for r_{ix}, i.e.,

$$f(r_{ix}=1|\phi_2,r_{iy}y_i)=\frac{\exp(\phi_{20}+\phi_{21}r_{iy}y_i)}{1+\exp(\phi_{20}+\phi_{21}r_{iy}y_i)}.$$

From (5.7), we obtain

$$f(y_i|\beta,\alpha,\phi,x_i,r_i=\mathbf{1})=\frac{\exp\left\{-\frac{(y_i-\beta_0-\beta_1 x_i)^2}{2}\right\}\left[\frac{\exp(\phi_{20}+\phi_{21}y_i)}{1+\exp(\phi_{20}+\phi_{21}y_i)}\right]}{\int\exp\left\{-\frac{(y_i-\beta_0-\beta_1 x_i)^2}{2}\right\}\left[\frac{\exp(\phi_{20}+\phi_{21}y_i)}{1+\exp(\phi_{20}+\phi_{21}y_i)}\right]dy_i}.$$

Thus, $E[y_i|\beta, \alpha, \phi, x_i, r_i = \mathbf{1}] \neq \beta_0 + \beta_1 x_i$. In this case, both $\widehat{\beta}_{cc}$ and $\widehat{\alpha}_{cc}$ may be biased. Again, an analytical derivation of the Fisher information matrix is not possible.

6 Simulation Studies and a Real Data Example

In this section, we present two detailed simulation studies and a real data example, demonstrating the various properties of the CC, CR, and AC methodology for analyzing MCAR and MAR response and/or covariate data in linear regression and logistic regression. In particular, we study efficiency and bias in the estimates for the three methods for these two types of regression models.

6.1 Simulation Study I: Normal Linear Regression Model with MCAR Response and Covariate Data

We consider a multiple linear regression model with an intercept, a completely observed covariate and two missing covariates. 5000 replicates with $n = 500$ subjects are considered. The response model is $y_i \sim N(\beta_0 + \beta_1 z_{i1} + \beta_2 x_{i1} + \beta_3 x_{i2}, 1)$, where z_i is simulated from Unif(0, 1), x_{i1} is simulated from $N(\alpha_{10} + \alpha_{11}z_{i1}, 1)$, x_{i2} is simulated from $N(\alpha_{20} + \alpha_{21}z_{i1} + \alpha_{22}x_{i1}, 1)$, and y_i, x_{i1} and x_{i2} are missing for some subjects. In each simulation, the sizes for each missing pattern, n_{11}, n_2, n_{23}, n_{24}, n_3, n_{42}, n_{43} and n_5, were varied in order to evaluate the various properties of the CC, CR, and AC methods. To better study the differences of the asymptotic variances of the estimates of the regression coefficients using the three methods, we calculated the variances in two ways: plugging in the true parameter values as well as plugging in the maximum likelihood estimates into the Fisher information matrix.

Table 3 gives the simulation results of the linear regression model with the variances evaluated at the true parameter values. We note here that the variance estimates decrease monotonically based on the three methods, CC, CR, and AC. In particular, Result 4.3 and Result 4.4 hold and Remark 1 – Remark 3 hold, but not Remark 4. We note that Remark 1 is only true for the regression coefficients of the missing covariates but not of the completely observed covariates. The information in Block B_{43} does improve the asymptotic variances of regression coefficients of the $z_{i0} = 1$ and z_{i1} when the AC method is used.

Table 4 gives the simulation results of the linear regression model with the variances evaluated at the maximum likelihood estimates (MLE's). The results show the gain in using the AC method over CR method, and using CR method over CC method. When when $n_{22} = 200$ and $n_{23} = n_{24} = 0$, the gain on the asymptotic variance of β_3 is small compared to the AC to CR method and the difference of the empirical variances is slight.

6.2 Simulation Study II: Logistic Regression Model with MAR Response and Covariate Data

A simulation with 1000 replicates was conducted to numerically compare the CC, CR and AC methods in a logistic regression model. The estimates using the full data (FD) before missing are also provided as a benchmark of other methods. In each simulation, we generated 500 binary samples from a logistic regression model $\text{logit}(P(y_i = 1)) = \beta_0 + \beta_1 z_{i1} + \beta_2 x_{i1}$, where z_{i1} was simulated from Unif(0, 1) and x_{i1} was simulated from a Bernoulli distribution with the success probability modeled as $\text{logit}(P(x_{i1} = 1)) = \alpha_0 + \alpha_1 z_{i1}$. The covariate z_{i1} is completely observed for all subjects, and x_{i1} and the response y_i are missing at random (MAR) for some subjects. The missing mechanisms for y_i and x_{i1} are $\text{logit}(P(r_{iy} = 1)) = \phi_{20} + \phi_{21} z_{i1}$ and $\text{logit}(P(r_{ix} = 1)) = \phi_{10} + \phi_{11} z_{i1} + \phi_{12} r_{iy} y_i$, where $r_{iy} = 1$ or $r_{ix} = 1$ if y_i or x_{i1} is observed, 0 otherwise. On average, 31.4% samples have completely observed covariate and response, 34.8% have missing covariate but observed response, 15.0% have observed covariate but missing response, and 18.8% have missing covariate and missing response.

Table 5 gives the simulation results of the logistic regression model. The AC method provides estimates with higher precision (smaller standard error) and lower mean square error (MSE) than the CR method for all the parameters. Both the CR and AC methods are uniformly better than the CC method in terms of bias, simulated standard error and MSE.

6.3 Analysis of Small Cell Lung Cancer Data

We consider a real dataset to compare the three analysis methods in terms of bias and efficiency. We consider a lung cancer dataset from a recent phase III clinical trial (Socinski et al., 2002) of non-small-cell lung cancer (NSCLC), which is the leading cause of cancer-related mortality. In the year 2001, among approximately 170,000 patients newly diagnosed, more than 90% died from NSCLC and approximately 35% of all new cases were stage IIIB/IV (malignant pleural effusion) the disease. A randomized, two-armed, multi-center trial was initiated in 1998 with the aim to determine the optimal duration of chemotherapy by comparing four cycles of therapy versus continuous therapy in advanced NSCLC. Patients were randomized to two treatment arms: four cycles of carboplatin at an area under the curve of 6 and paclitaxel 200 mg/m^2 every 21 days (arm A), or continuous treatment with carboplatin/paclitaxel until progression (arm B). At progression, all patients on both arms received second-line weekly paclitaxel at 80 mg/m^2. One of the primary endpoints was quality of life (QOL). There were $n = 230$ patients in this dataset. The response variable considered in this analysis is the quality of life (QOL) factg score. The covariates included in the model were treatment (trt, 0=arm A, 1=arm B), gender (0=female, 1=male), Histology (hist, 0=Non-Squamous, 1=Squamous), age at entry in years, highest grade toxicity (recorded by cycle) (apex, 0 if highest grade toxicity =0 and 1 if highest grade toxicity > 0), and recovery status (recov, 0 if recovered and 1 otherwise). For these six covariates, apex and recov had missing information and trt, gender, hist, and age were completely observed for all cases. In this population, 63% of the patients were male, and the age at entry ranged from 32 to 82 with a mean of 62. The missing data fractions were 28% in apex, 54% in recov, and 35% in factg. There was a total missing data fraction of 74% on apex, recov, and factg.

We use a linear regression to model the response variable, factg, as

$$\text{factg}_i = \beta_0 + \beta_1 \text{trt}_i + \beta_2 \text{gender}_i + \beta_3 \text{hist}_i + \beta_4 \text{age}_i + + \beta_5 \text{apex}_i + \beta_6 \text{recov}_i + \varepsilon_i.$$

We consider two models for the missing covariates recov and apex as follows.

Model M1—

$$\text{logit(apex}_i=1)=\alpha_{10}+\alpha_{11}\text{trt}_i+\alpha_{12}\text{gender}_i+\alpha_{13}\text{hist}_i+\alpha_{14}\text{age}_i,$$
$$\text{logit(recov}_i=1|\text{apex}_i=1)=\alpha_{20}+\alpha_{21}\text{trt}_i+\alpha_{22}\text{gender}_i+\alpha_{23}\text{hist}_i+\alpha_{24}\text{age}_i+\alpha_{25}\text{apex}_i.$$

Model M2—

$$\text{logit(recov}_i=1)=\alpha_{10}+\alpha_{11}\text{trt}_i+\alpha_{12}\text{gender}_i+\alpha_{13}\text{hist}_i+\alpha_{14}\text{age}_i,$$
$$\text{logit(apex}_i=1|\text{recov}_i)=\alpha_{20}+\alpha_{21}\text{trt}_i+\alpha_{22}\text{gender}_i+\alpha_{23}\text{hist}_i+\alpha_{24}\text{age}_i+\alpha_{25}\text{recov}_i.$$

Table 6 shows the results for the CC, CR and AC methods discussed in Section 4. We assume that the missing data are MAR so that a missing data mechanism need not be considered in the estimation scheme for $\boldsymbol{\beta}$. As shown in the table, the overall conclusions are the same for the CR and AC methods, as these two methods yield similar p-values for the various regression coefficients. significance level. However, the CR and AC methods yield more significant p-values than the CC analysis, especially for the age effect. Table 6 also shows that the estimates of β and the standard errors of the estimated regression coefficients are quite similar for model M1 and model M2, indicating robustness of estimates to the the choice of covariate distribution. The EM algorithm was implemented for computing all maximum likelihood estimates. The convergence criterion for the EM algorithm was that the squared distance between the kth and $(k + 10)^{th}$ iterations was less than 10^{-7}. The EM algorithm required 25 iterations to converge under both model M1 and model M2.

Although we have assumed that data are MAR and a missing data mechanism need not be modeled, it is of some interest if we could determine the best fitting MAR missing data mechanism, so that we could at least (though somewhat ad-hoc) determine whether the missing data are MAR or MCAR. Towards this goal, we posited several different MAR and MCAR missing data mechanisms, and using the complete cases to fit these models as logistic regression in SAS. We then computed the log-likelihood statistic to determine the best fitting missing data mechanism. We considered 5 missing data mechanism, two of them are MCAR, and the other three are MAR. Let $r_{i,\text{factg}}$, $r_{i,\text{apex}}$, and $r_{i,\text{recov}}$ denote the missing data indicators for factg, apex, and recov, respectively. To determine the final log-likelihood statistic value, we added the three contributions from the three parts of the binary regression models for $r_{i,\text{factg}}$, $r_{i,\text{apex}}$, and $r_{i,\text{recov}}$. Two MCAR models are $[r_{i,\text{factg}}][r_{i,\text{apex}}][r_{i,\text{recov}}]$ (MCAR1) and $[r_{i,\text{factg}}|r_{i,\text{apex}},$ $r_{i,\text{recov}}][r_{i,\text{apex}}][r_{i,\text{recov}}|r_{i,\text{apex}}]$ (MCAR2), where, for example, $[r_{i,\text{factg}}]$ denotes a logistic regression model with intercept only and $[r_{i,\text{factg}}|r_{i,\text{apex}}, r_{i,\text{recov}}]$ is a logistic regression model with intercept and covariates $r_{i,\text{apex}}$ and $r_{i,\text{recov}}$. Let $x_{i,obs} = (\text{trt}_i, \text{gender}_i, \text{hist}_i, \text{age}_i)$. Three MAR models include $[r_{i,\text{factg}}|x_{i,obs}, \text{apex}_i r_{i,\text{apex}}, \text{recov}_i r_{i,\text{recov}}][r_{i,\text{apex}}|x_{i,obs}][r_{i,\text{recov}}|x_{i,obs},$ $\text{apex}_i r_{i,\text{apex}}]$ (MAR1), $[r_{i,\text{factg}}|x_{i,obs}, r_{i,\text{apex}}, r_{i,\text{recov}}][r_{i,\text{apex}}|x_{i,obs}][r_{i,\text{recov}}|x_{i,obs}, r_{i,\text{apex}}]$ (MAR2), and $[r_{i,\text{factg}}|\text{trt}_i, \text{gender}_i, r_{i,\text{apex}}, r_{i,\text{recov}}][r_{i,\text{apex}}|\text{trt}_i, \text{gender}_i][r_{i,\text{recov}}|\text{trt}_i, \text{gender}_i, r_{i,\text{apex}}]$ (MAR3). For the lung cancer data, the log-likelihood statistics, $-2\log(\text{likelihood})$, are 889.4, 873.5, 868.5, 856.6, and 860.5 under models MCAR1, MCAR2, MAR1, MAR2, and MAR3, respectively. We see from these results that the best fitting model is MAR2 as the missing data mechanism suggesting that the missing data are missing at random.

7 Discussion

We have given several results regarding bias and efficiency of estimates in missing (x, y) regression problems, and have shown that the AC analysis provides the most efficient estimates and the least biased estimates in the MAR setting. The results derived in Section 4, Section 5, and Section 6 are new and important and shed light on the bias and efficiency of estimates in regression problems with MCAR or MAR responses and/or covariates. In Section 4.1 and 4.2,

the variances are assumed to be known. This assumption can be relaxed. The asymptotic variance and covariance matrix of the MLE's under each analysis method for the simple linear regression model with unknown variances is derived in Appendix B. With the known variances, as shown in Result 4.2, the AC analysis does not improve the efficiency of the MLE for β_1 over the CR analysis. When the variances are unknown, it is interesting to see, from Appendix B, that the AC analysis improves the efficiency of the MLE for $^2_1$ over both the CC and CR analsyese. Thus, the AC analysis becomes even more important in this case. However, the derivation of the asymptotic variance and covariance matrix of the MLE's under each analysis method for the multiple linear regression model with unknown variances becomes very lengthy and hence, detailed derivations are omitted for brevity.

Finally, we mention that we have assumed throughout that *jointly*, (x_i, y_i) are iid. This is by far the most common approach in regression settings with missing covariate and/or response data. We note here, however, that since inference typically focuses on the parameters of $[y_i/x_i]$, the y_i's conditional on the x_i's are *not* iid, but rather only independent. This development is still quite general since is covers settings such as the linear model and generalized linear models with MAR covariate and/or response data. Future work involves examination of the proposed methods for dependent responses, including dynamic linear models, models for longitudinal data, and generalized linear mixed models. Such theoretical investigations are currently being examined. The initial investigation taken here is the first of its kind, and should lead to fruitful results for other types of models.

Acknowledgements

Appendix A

Computational Development

We describe the model fitting and computational procedures for each of the analysis methods. For CC, the MLE of θ can be obtained by standard statistical software such as SAS. Here, we consider only for AC as the computation of the MLE's under CR is similar to and even easier than AC.

We first consider the case where all $x_{i,mis}$'s are categorical. In this case, we use the EM algorithm via the method of weights proposed by Ibrahim (1990). Let $\theta^{(t)} = (\beta^{(t)}, \phi^{(t)}, \alpha^{(t)})$ denote the value of θ at the tth iteration of EM algorithm.

The E-step at the $(t + 1)^{st}$ iteration can be written as

$$
\begin{aligned}
Q(\theta|\theta^{(t)}) &= \sum_{i:\, y_i \text{ observed},\, M_i=\theta} [\log f(y_i|x_i,\beta,\phi) + \log f(x_i|\alpha)] \\
&+ \sum_{i:\, y_i \text{ observed},\, M_i \neq \theta x_{i,\mathrm{mis}}(j)} \sum w_{ij,(t)} [\log f(y_i|x_i,(j),\beta,\phi) + \log f(x_i(j)|\alpha)] \\
&+ \sum_{i:\, y_i \text{ missing},\, M_i=\theta} \log f(x_i|\alpha) \\
&+ \sum_{i:\, y_i \text{ missing},\, M_i \neq \theta, M_i \neq \Omega x_{i,\mathrm{mis}}(j)} \sum w_{ij,(t)} \log f(x_i(j)|\alpha) \\
&= Q^{(1)}(\beta,\phi|\theta^{(t)}) + Q^{(2)}(\alpha|\theta^{(t)}),
\end{aligned}
\tag{A.1}
$$

where $x_i(j) = (x'_{i,\mathrm{obs}}, x'_{i,\mathrm{mis}}(j))'$,

$$
\begin{aligned}
Q^{(1)}(\beta,\phi|\theta^{(t)}) &= \sum_{i:\, y_i \text{ observed},\, M_i=\theta} \log f(y_i|x_i,\beta,\phi) \\
&+ \sum_{i:\, y_i \text{ observed},\, M_i\neq\theta}\sum_{x_{i,\mathrm{mis}}(j)} w_{ij,(t)} \log f(y_i|x_i(j),\beta,\phi)
\end{aligned}
\tag{A.2}
$$

and

$$
\begin{aligned}
Q^{(2)}(\alpha|\theta^{(t)}) &= \sum_{i:\, y_i \text{ observed or missing},\, M_i=\theta} \log f(x_i|\alpha) \\
&+ \sum_{i:\, y_i \text{ observed},\, M_i\neq\theta}\sum_{x_{i,\mathrm{mis}}(j)} w_{ij,(t)} \log f(x_i(j)|\alpha) \\
&+ \sum_{i:\, y_i \text{ missing},\, M_i\neq\theta,\, M_i\neq\Omega}\sum_{x_{i,\mathrm{mis}}(j)} w_{ij,(t)} \log f(x_i(j)|\alpha).
\end{aligned}
\tag{A.3}
$$

The inner sum extends over all of the possible values of the missing components of the covariate vector, with j indexing the distinct covariate patterns for subject i.

The weights, $w_{ij,(t)}$, are the conditional probabilities corresponding to $[\boldsymbol{x}_{i,mis}|\boldsymbol{x}_{i,obs}, y_i\, \boldsymbol{\theta}]$ or $[\boldsymbol{x}_{i,mis}|\boldsymbol{x}_{i,obs},\, \boldsymbol{\alpha}]$ and are given by

$$
\begin{aligned}
w_{ij,(t)} &= f(x_{i,\mathrm{mis}}(j)|x_{i,\mathrm{obs}},y_i,\theta^{(t)}) \\
&= \frac{f(y_i|x_{i,\mathrm{obs}},x_{i,\mathrm{mis}}(j),\beta^{(t)},\phi^{(t)})f(x_{i,\mathrm{obs}},x_{i,\mathrm{mis}}(j)|\alpha^{(t)})}{\sum_{x_{i,\mathrm{mis}}(j)} f(y_i|x_{i,\mathrm{obs}},x_{i,\mathrm{mis}}(j),\beta^{(t)},\phi^{(t)})f(x_{i,\mathrm{obs}},x_{i,\mathrm{mis}}(j)|\alpha^{(t)})}
\end{aligned}
$$

or

$$
w_{ij,(t)} = f(x_{i,\mathrm{mis}}(j)|x_{i,\mathrm{obs}},\alpha^{(t)}) = \frac{f(x_{i,\mathrm{obs}},x_{i,\mathrm{mis}}(j)|\alpha^{(t)})}{\sum_{x_{i,\mathrm{mis}}(j)} f(x_{i,\mathrm{obs}},x_{i,\mathrm{mis}}(j)|\alpha^{(t)})}.
$$

The M-step at the $(t + 1)^{st}$ iteration proceeds as follows. We first compute

$$
(\beta^{(t+1)},\phi^{(t+1)}) = \underset{\beta,\phi}{\arg\max}\, Q^{(1)}(\beta,\phi\,|\,\theta^{(t)})
$$

and

$$
\alpha^{(t+1)} = \underset{\alpha}{\arg\max}\, Q^{(2)}(\alpha\,|\,\theta^{(t)}).
$$

When we use the saturated model for $\boldsymbol{x}_{i,mis}$, we have $\boldsymbol{\alpha} = (\alpha_{(j)})$, where $\alpha_{(j)}$ denotes the probability of the jth missing pattern. In this case, we update

$$\alpha_{(j)}^{(t+1)} = \frac{\sum_{i=1}^{n} w_{ij,(t)}}{n}$$

where when x_i is completely observed, $w_{ij,(t)} = 1$ if $x_i(j) = x_i$ and $w_{ij,(t)} = 0$ if $x_i(j) \neq x_i$.

Let $\hat{\theta}$ denote the estimate of θ at EM convergence. We use Louis's method (Louis, 1982) to compute the estimated observed information matrix of θ based on the observed data. Write the matrix of second derivatives of $Q(\theta|\theta^{(t)})$ as

$$\ddot{Q}(\theta|\theta^{(t)}) = \begin{pmatrix} \frac{\partial^2}{\partial\gamma\partial\gamma'} Q^{(1)}(\beta,\phi \mid \theta^{(t)}) & 0 \\ 0 & \frac{\partial^2}{\partial\alpha\partial\alpha'} Q^{(2)}(\alpha \mid \theta^{(t)}) \end{pmatrix},$$

where $\gamma = (\beta', \phi')'$,

$$\frac{\partial^2}{\partial\gamma\partial\gamma'} Q^{(1)}(\beta,\phi \mid \theta^{(t)}) = \sum_{i:\, y_i \text{ observed, } M_i=\theta} \frac{\partial^2}{\partial\gamma\partial\gamma'} \log f(y_i|x_i,\beta,\phi)$$
$$+ \sum_{i:\, y_i \text{ observed, } M_i\neq\theta} \frac{\partial^2}{\partial\gamma\partial\gamma'} \sum_{x_{i,\text{mis}}(j)} w_{ij,(t)} \log f(y_i|x_i(j),\beta,\phi),$$

and

$$\frac{\partial^2}{\partial\alpha\partial\alpha'} Q^{(2)}(\alpha \mid \theta^{(t)}) = \sum_{i:\, y_i \text{ observed or missing, } M_i=\theta} \frac{\partial}{\partial^2\alpha\partial\alpha'} \log f(x_i|\alpha)$$
$$+ \sum_{i:\, y_i \text{ observed, } M_i\neq\theta} \sum_{x_{i,\text{mis}}(j)} w_{ij,(t)} \frac{\partial^2}{\partial\alpha\partial\alpha'} \log f(x_i(j)|\alpha)$$
$$+ \sum_{i:\, y_i \text{ missing, } M_i\neq\theta, M_i\neq\Omega} \sum_{x_{i,\text{mis}}(j)} w_{ij,(t)} \frac{\partial^2}{\partial\alpha\partial\alpha'} \log f(x_i(j)|\alpha).$$

Write the gradient vector of $Q_i(\theta|\theta^{(t)})$ for the ith observation as

$$\dot{Q}_i(\theta|\theta^{(t)})$$
$$= \begin{cases} \frac{\partial}{\partial\theta} \log[\, f(y_i|x_i,\beta,\phi)f(x_i|\alpha)] & \text{if } y_i \text{ observed, } M_i=\theta \\ \sum_{x_{i,\text{mis}}(j)} w_{ij,(t)} \frac{\partial}{\partial\theta} \log[\, f(y_i|x_i(j),\beta,\phi)f(x_i(j)|\alpha)] & \text{if } y_i \text{ observed, } M_i \neq \theta \\ \frac{\partial}{\partial\theta} \log f(x_i|\alpha) & \text{if } y_i \text{ missing, } M_i=\theta \\ \sum_{x_{i,\text{mis}}(j)} w_{ij,(t)} \frac{\partial}{\partial\theta} \log f(x_i(j)|\alpha) & \text{if } y_i \text{ missing, } M_i \neq \theta, M_i \neq \Omega. \end{cases}$$

In addition, write the complete data score vector $S_i(\theta|x_i, y_i)$ as

$$S_i(\theta|x_i,y_i) = \begin{cases} \frac{\partial}{\partial\theta} \log [\, f(y_i|x_i,\beta,\phi)f(x_i|\alpha)] & \text{if } y_i \text{ observed} \\ \frac{\partial}{\partial\theta} \log f(x_i|\alpha) & \text{if } y_i \text{ missing.} \end{cases}$$

Then, the estimated observed information matrix of $\hat{\theta}$ is given by

$$I(\widehat{\theta})= \ -\ddot{Q}(\widehat{\theta|\theta})$$

$$-\left\{\left[\sum_{i:M_i\neq\theta}\sum_{x_{i,\mathrm{mis}(j)}}w_{ij,(t)}S_i(\widehat{\theta}|x_i(j),y_i)S_i(\widehat{\theta}|x_i(j),y_i)'\right] - \sum_{i:M_i\neq\theta}\dot{Q}_i(\widehat{\theta|\theta})\dot{Q}_i(\widehat{\theta|\theta})'\right\},$$

(A.4)

where the weights, $w_{ij,(t)}$, are computed at EM convergence. Thus, the estimate of the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}$ is $[I(\hat{\boldsymbol{\theta}})]^{-1}$.

When missing covariates are continuous or mixed Continuous and categorical, a Monte Carlo EM (MCEM) algorithm is required. The implementation of the MCEM is similar to the EM algorithm for the categorical missing covariates, and is developed in detail in Ibrahim, Lipsitz and Chen (1999), and Ibrahim, Chen, and Lipsitz (1999). Specifically, we replace the weight average in (A.2), (A.3) and (A.4) by a Monte Carlo average. For example, in the E-Step, for missing covariates in the Block B_2, we take an MCMC sample of size

$$m_i^{(t)}, x_{i,\mathrm{mis}}^{(t1)}, x_{i,\mathrm{mis}}^{(t2)}, \ldots, x_{i,\mathrm{mis}}^{(tm^{(t)})}, \text{from}$$

$$f(x_{i,\mathrm{mis}}|x_{i,\mathrm{obs}},y_i,\theta^{(t)}) \propto f(y_i|x_{i,\mathrm{obs}},x_{i,\mathrm{mis}},\beta^{(t)},\phi^{(t)})f(x_{i,\mathrm{obs}},x_{i,\mathrm{mis}}|\alpha^{(t)}).$$

Then, we compute

$$Q^{(1)}(\beta,\phi\,|\,\theta^{(t)})= \sum_{i:\,y_i\text{ observed, }M_i=\theta} \log f(y_i|x_i,\beta,\phi)$$

$$+ \sum_{i:\,y_i\text{ observed, }M_i\neq\theta} \frac{1}{m^{(t)}}\sum_{l=1}^{m^{(t)}}\log f(y_i|x_i^{(tl)},\beta,\phi),$$

$x_i^{(tl)}=(x_{i,\mathrm{obs}}', x_{i,\mathrm{mis}}^{(tl)'})'$. We then take $m^{(t+1)} = m^{(t)} + \Delta m$, where $\Delta m > 0$. In such a way, the MCEM algorithm requires much less computational time, as a large $m^{(t)}$ is not needed in early iterations of the algorithm.

Appendix B

Simple Linear Regression Model with Unknown Variances

We consider a simple normal regression model with a single covariate and unknown variances here. In this case, we have

$$f(y_i|x_i,\beta,\sigma)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(y_i-\beta_0-\beta_1 x_i)^2}{2\sigma^2}\right\} \text{ and } f(x_i|\alpha,\tau)=\frac{1}{\sqrt{2\pi\tau^2}}\exp\left\{-\frac{(x_i-\alpha)^2}{2\tau^2}\right\}.$$

Write $\boldsymbol{\theta} = (\boldsymbol{\beta}', \sigma^2, \alpha, \tau^2)'$. Let $n_j = \#(B_j)$ be the cardinality of B_j for $j = 1, 2, 3$ and $n = n_1 + n_2 + n_3$. For the CC analysis, we have

$$l_{cc}(\theta)= \sum_{i:\,y_i\text{ observed, }M_i=\theta}\left[-\log(2\pi\sqrt{\sigma^2\tau^2}) - \frac{(y_i-\beta_0-\beta_1 x_i)^2}{2\sigma^2} - \frac{(x_i-\alpha)^2}{2\tau^2}\right]$$

and the Fisher information matrix is given by

$$I_{cc}(\theta) = - E\left[\frac{\partial^2}{\partial\theta\partial\theta'} l_{cc}(\theta)\right] = n_1 \begin{pmatrix} 1/\sigma^2 & \alpha/\sigma^2 & 0 & 0 & 0 \\ \alpha/\sigma^2 & (\tau^2+\alpha^2)/\sigma^2 & 0 & 0 & 0 \\ 0 & 0 & 1/(2\sigma^4) & 0 & 0 \\ 0 & 0 & 0 & 1/\tau^2 & 0 \\ 0 & 0 & 0 & 0 & 1/(2\tau^4) \end{pmatrix}.$$

For the CR analysis, we have

$$l_{cr}(\theta) = l_{cc}(\theta) + \sum_{i\in B_2}\left\{-\log\sqrt{2\pi} - \frac{\log(\upsilon)}{2} - \frac{(y_i - \beta_0 - \beta_1\alpha)^2}{2\upsilon}\right\},$$

where $\upsilon = \sigma^2 + \beta_1^2\tau^2$. After some messy algebra, we obtain the Fisher information matrix given by

$$I_{cr}(\theta) = - E\left[\frac{\partial^2}{\partial\theta\partial\theta'} l_{cr}(\theta)\right] = I_{cc}(\theta) + \frac{n_2}{\upsilon^2}\begin{pmatrix} \upsilon & \alpha\upsilon & 0 & \beta_1\upsilon & 0 \\ \alpha\upsilon & \alpha^2\upsilon + 2\beta_1^2\tau^4 & \beta_1\tau^2 & \alpha\beta_1\upsilon & \beta_1^3\tau^2 \\ 0 & \beta_1\tau^2 & 1/2 & 0 & \beta_1^2/2 \\ \beta_1\upsilon & \alpha\beta_1\upsilon & 0 & \beta_1^2\upsilon & 0 \\ 0 & \beta_1^3\tau^2 & \beta_1^2/2 & 0 & \beta_1^4 \end{pmatrix}.$$

For the AC analysis, the log-likelihood function is given by

$$l_{ac}(\theta) = l_{cr}(\theta) + \sum_{i:\, y_i\text{ missing},\, M_i=0} \log f(x_i|\alpha) = l_{cr}(\theta) + \sum_{i\in B_3}\left[-\log\sqrt{2\pi\tau^2} - \frac{(x_i - \alpha)^2}{2\tau^2}\right].$$

The corresponding Fisher information matrix is given by

$$I_{ac}(\theta) = - E\left[\frac{\partial^2}{\partial\theta\partial\theta'} l_{ac}(\theta)\right] = I_{cr}(\theta) + n_3\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/\tau^2 & 0 \\ 0 & 0 & 0 & 0 & 1/(2\tau^4) \end{pmatrix}.$$

Based on either the determinant or the trace of Fisher information matrix, AC yields most gain in information over both CR and CC, and CR gains more information than CC. Specifically, we have

$$
\begin{aligned}
&\det(I_{ac}) - \det(I_{cr})\\
=\ & \frac{n_1 n_3}{4\upsilon^5\sigma^8\tau^4}\Big[n_1^2\upsilon^5(2n_1+3n_2+n_3) + n_1n_2\sigma^2\upsilon^4(n_1+2n_2+2n_3) + n_2\sigma^2\upsilon^3(n_2n_3\sigma^2 + n_1n_2\beta_1^2\tau^2\\
& +n_1n_3\beta_1^2\tau^2) + n_2^2\beta_1^2\sigma^2\tau^2\upsilon^2(n_3\sigma^2 + 2n_1\beta_1^2\tau^2) + n_2^2\beta_1^4\sigma^2\tau^4\upsilon(n_2\sigma^2 + n_1\beta_1^2\tau^2) + n_2^3\beta_1^6\sigma^4\tau^6\Big]\\
>\ & 0,
\end{aligned}
$$

$$\det(I_{cr}) - \det(I_{cc})$$
$$= \frac{n_1^2 n_2}{4 v^4 \sigma^8 \tau^4} \left[n_1 v^4 (2n_1+n_2) + n_1 \beta_1^4 \tau^4 v^2 (n_1+n_2) + n_2 \beta_1^4 \sigma^2 \tau^4 (n_1+n_2)(v+\beta_1^2 \tau^2) \right] > 0,$$

$$\mathrm{tr}(I_{ac}) - \mathrm{tr}(I_{cr}) = \frac{n_3(1+2\tau^2)}{2\tau^4} > 0,$$

and

$$\mathrm{tr}(I_{cr}) - \mathrm{tr}(I_{cc}) = \frac{n_2}{2v^2} \left[2v(1+\beta_1^2+\alpha^2) + 4\beta_1^2 \tau^4 + 2\beta_1^4 + 1 \right] > 0.$$

CR leads to smaller asymptotic variances for all parameters than CC. Specifically, we have

$$\begin{aligned}
\mathrm{Var}(\widehat{\beta}_{1,cc}) - \mathrm{Var}(\widehat{\beta}_{1,cr}) &= 2n_2 \beta_1^2 \sigma^4 (n_2 \beta_1^4 \tau^4 + n_1 v^2)/a_1 > 0, \\
\mathrm{Var}(\widehat{\sigma}_{cc}^2) - \mathrm{Var}(\widehat{\sigma}_{cr}^2) &= 2n_2 \sigma^8 (n_2 \beta_1^4 \tau^4 + n_1 v^2)/a_1 > 0, \\
\mathrm{Var}(\widehat{\alpha}_{cc}) - \mathrm{Var}(\widehat{\alpha}_{cr}) &= n_2 \beta_1^2 \tau^4/n_1(n_1+n_2)v > 0, \\
\mathrm{Var}(\widehat{\tau}_{cc}^2) - \mathrm{Var}(\widehat{\tau}_{cr}^2) &= 2n_2 \beta_1^4 \tau^8 [2n_1 v^2 + n_2 \sigma^2 (v+\beta_1^2 \tau^2)]/a_1 > 0,
\end{aligned}$$

where $a_1 = n_1(n_1^2 v^4 + n_2^2 \beta_1^4 \sigma^2 \tau^4 (v+\beta_1^2 \tau^2) + n_1 n_2 v^2 (v^2 + \beta_1^4 \tau^4))$.

In addition, AC improves the asymptotic variances over CR. Specifically, we have

$$\begin{aligned}
\mathrm{Var}(\widehat{\beta}_{1,cr}) - \mathrm{Var}(\widehat{\beta}_{1,ac}) &= 2v^4 n_2^2 n_3 \beta_1^6 \sigma^4 \tau^4/a_2 a_3 > 0, \\
\mathrm{Var}(\widehat{\sigma}_{cr}^2) - \mathrm{Var}(\widehat{\sigma}_{ac}^2) &= 2v^4 n_2^2 n_3 \beta_1^4 \sigma^8 \tau^4/a_2 a_3 > 0, \\
\mathrm{Var}(\widehat{\alpha}_{cr}) - \mathrm{Var}(\widehat{\alpha}_{ac}) &= n_3 \tau^2 (n_2 \sigma^2 + n_1 v)/n_1(n_1+n_2)(n_2 n_3 \sigma^2 + n_1(n_1+n_2+n_3)v)v > 0, \\
\mathrm{Var}(\widehat{\tau}_{cr}^2) - \mathrm{Var}(\widehat{\tau}_{ac}^2) &= 2n_3 \tau^4 v^4 [n_1 v^2 + n_2 \sigma^2 (v+\beta_1^2 \tau^2)]^2/a_2 a_3 > 0,
\end{aligned}$$

where

$a_2 = v^4 n_1(n_1+n_2) + n_2 \beta_1^4 \sigma^4 v(n_1 v + n_2 \sigma^2) + n_2^2 \beta_1^6 \sigma^2 \tau^6$, and $a_3 = n_1(n_1+n_2+n_3)v^4 + n_2 \sigma^2 (v+\beta_1^2 \tau^2)(n_3 v^2 + n_2 \beta_1^4 \tau^4) + n_1 n_2 \beta_1^4 \tau^4 v^2$

.

References

Baker SG, Laird NM. Regression Analysis for Categorical Variables with Outcome Subject to Nonignorable Nonresponse. Journal of the American Statistical Association 1988;83:62–69.

Chen Q, Ibrahim JG. Semiparametric Models for Missing Covariate and Response Data in Regression Models. Biometrics 2006;62:177–184. [PubMed: 16542244]

Chen M-H, Ibrahim JG, Shao Q-M. Propriety of the Posterior Distribution and Existence of the Maximum Likelihood Estimator for Regression Models with Covariates Missing at Random. Journal of the American Statistical Association 2004;99:421–438.

Herring AH, Ibrahim JG. Likelihood-based Methods for Missing Covariates in the Cox Proportional Hazards Model. Journal of the American Statistical Association 2001;96:292–302.

Horton NJ, Laird NM. Maximum Likelihood Analysis of Generalized Linear Models with Missing Covariates. Statistical Methods in Medical Research 1999;8:37–50. [PubMed: 10347859]

Ibrahim JG. Incomplete data in generalized linear models. Journal of the American Statistical Association 1990;85:765–769.

Ibrahim JG, Chen M-H, Lipsitz SR. Monte Carlo EM for Missing Covariates in Parametric Regression Models. Biometrics 1999;55:591–596. [PubMed: 11318219]

Ibrahim JG, Chen M-H, Lipsitz SR. Missing Responses in Generalized Linear Mixed Models When the Missing Data Mechanism Is Nonignorable. Biometrika 2001;88:551–564.

Ibrahim JG, Chen M-H, Lipsitz SR, Herring AH. Missing Data Methods for Generalized Linear Models: A Comparative Review. Journal of the American Statistical Association 2005;100:332–346.

Ibrahim JG, Lipsitz SR, Chen M-H. Missing Covariates in Generalized Linear Models When the Missing Data Mechanism is Nonignorable. Journal of the Royal Statistical Society, Series B 1999;61:173–190.

Lipsitz SR, Ibrahim JG, Zhao LP. A New Weighted Estimating Equation for Missing Covariate Data with Properties Similar to Maximum Likelihood. Journal of the American Statistical Association 1999;94:1147–1160.

Little RJA. Regression with Missing X's: A Review. Journal of the American Statistical Association 1992;87:1227–1237.

Louis TA. Finding the Observed Information Matrix When Using the EM Algorithm. Journal of the Royal Statistical Society, Series B 1982;44:226–233.

Robins JM, Rotnitzky A. Semiparametric Efficiency in Multivariate Regression Models with Missing Data. Journal of the American Statistical Association 1995;90:122–129.

Socinski MA, Schell MJ, Peterman A, Bakri K, Yates S, Gitten R, Unger P, Lee J, Lee Ji, Tynan M, Moore M, Kies M. Phase III Trial Comparing Defined Duration of Therapy Versus Continuous Therapy Followed by Second-LineTherapy in Advanced-Stage IIIB/IV Non-Small-Cell Lung Cancer. Journal of Clinical Oncology 2002;20:1335–1343. [PubMed: 11870177]

Stubbendick AL, Ibrahim JG. Maximum Likelihood Methods for Nonignorable Responses and Covariates in Random Effects Models. Biometrics 2003;59:1140–1150. [PubMed: 14969495]

Stubbendick AL, Ibrahim JG. Likelihood-based Inference with Nonignorable Missing Responses and Covariates in Models for Discrete Longitudinal Data. Statistica Sinica 2006;16:1143–1167.

Tang G, Little RJA, Raghunathan TE. Analysis of Multivariate Missing Data with Nonignorable Nonresponse. Biometrika 2003;90:747–764.
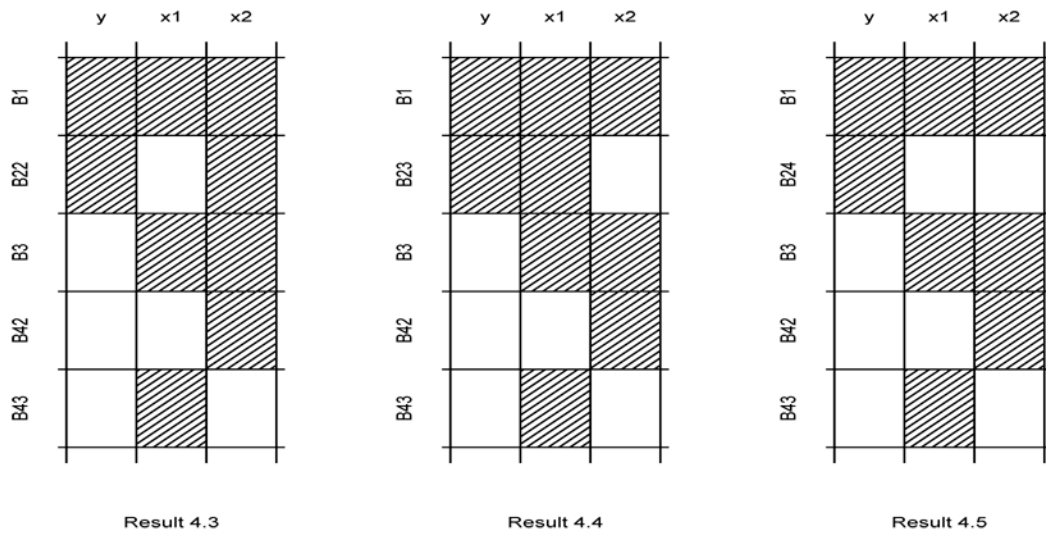
Figure 1. Missing Patterns
Note: Blank block stands for missing, marked block stands for observed.

Table 1

The Data Structure with Various Missing Patterns

Pattern (Block)	Response	Covariates
	y_i	x_i
B_1	observed	observed ($M_i = \theta$)
B_2	Observed	partially or all missing ($M_i \neq \theta$ or $M_i = \Omega$)
B_3	missing	observed ($M_i = \theta$)
B_4	missing	only partially missing ($M_i \neq \theta$ and $M_i \neq \Omega$)
B_5	missing	completely missing ($M_i = \Omega$)

Table 2

Monotonic Properties for CR versus CC and AC versus CR

Missing Pattern	Parameter	n_1	n_{22}	n_{23}	n_{24}
Result 4.3	$\mathrm{Var}(\hat{\beta}_{1,cc}) - \mathrm{Var}(\hat{\beta}_{1,cr})$	NM	NM		
$n_{23} = n_{24} = 0$	$\mathrm{Var}(\hat{\beta}_{2,cc}) - \mathrm{Var}(\hat{\beta}_{2,cr})$	NM	NM		
	$\mathrm{Var}(\hat{\beta}_{1,cr}) - \mathrm{Var}(\hat{\beta}_{1,ac})$	NM	NM		
	$\mathrm{Var}(\hat{\beta}_{2,cr}) - \mathrm{Var}(\hat{\beta}_{2,ac})$	NM	NM		
	$\mathrm{Var}(\hat{\beta}_{1,cc}) - \mathrm{Var}(\hat{\beta}_{1,cr})$	↗		NM	
Result 4.4	$\mathrm{Var}(\hat{\beta}_{2,cc}) - \mathrm{Var}(\hat{\beta}_{2,cr})$	↗		↗	
$n_{22} = n_{24} = 0$	$\mathrm{Var}(\hat{\beta}_{1,cr}) - \mathrm{Var}(\hat{\beta}_{1,ac})$	↗		↗	
	$\mathrm{Var}(\hat{\beta}_{2,cr}) - \mathrm{Var}(\hat{\beta}_{2,ac}) = 0$				↗
Result 4.5	$\mathrm{Var}(\hat{\beta}_{1,cc}) - \mathrm{Var}(\hat{\beta}_{1,cr})$	↗			↗
$n_{22} = n_{23} = 0$	$\mathrm{Var}(\hat{\beta}_{2,cc}) - \mathrm{Var}(\hat{\beta}_{2,cr})$	↗			↗
	$\mathrm{Var}(\hat{\beta}_{1,cr}) - \mathrm{Var}(\hat{\beta}_{1,ac})$	↗			↗
	$\mathrm{Var}(\hat{\beta}_{2,cr}) - \mathrm{Var}(\hat{\beta}_{2,ac})$	↗			↗

Table 3

Variance Comparison with True Parameters Plugged-in for Linear Regression

$(n_1, n_{22}, n_{23}, n_{24}, n_3, n_{42}, n_{43}, n_5)$	Para	CC (V_{CC})	CR (V_{CC}-V_{CR})	AC (V_{CR}-V_{AC})
	$\beta_0 = 1.0$	1.224×10^{-1}	6.111×10^{-2}	1.451×10^{-2}
	$\beta_1 = 1.0$	2.895×10^{-1}	1.420×10^{-1}	3.481×10^{-2}
	$\beta_2 = 2.0$	1.000×10^{-1}	6.419×10^{-2}	1.117×10^{-3}
(50, 200, 0, 0, 100, 100, 50, 0)	$\beta_3 = 3.0$	2.000×10^{-2}	1.336×10^{-2}	8.718×10^{-6}
	$\beta_0 = 1.0$	1.224×10^{-1}	6.111×10^{-2}	1.451×10^{-2}
	$\beta_1 = 1.0$	2.895×10^{-1}	1.420×10^{-1}	3.481×10^{-2}
	$\beta_2 = 2.0$	1.000×10^{-1}	6.419×10^{-2}	1.117×10^{-3}
(50, 200, 0, 0, 100, 100, 50, 0)	$\beta_3 = 2.5$	2.000×10^{-2}	1.336×10^{-2}	8.718×10^{-6}
	$\beta_0 = 1.0$	1.224×10^{-1}	6.111×10^{-2}	1.424×10^{-2}
	$\beta_1 = 1.0$	2.895×10^{-1}	1.420×10^{-1}	3.412×10^{-2}
	$\beta_2 = 2.0$	1.000×10^{-1}	6.419×10^{-2}	1.117×10^{-3}
(50, 200, 0, 0, 100, 100, 20, 30)	$\beta_3 = 3.0$	2.000×10^{-2}	1.336×10^{-2}	8.718×10^{-6}
	$\beta_0 = 1.0$	1.224×10^{-1}	1.660×10^{-2}	8.930×10^{-3}
	$\beta_1 = 1.0$	2.895×10^{-1}	3.001×10^{-2}	2.380×10^{-2}
	$\beta_2 = 2.0$	1.000×10^{-1}	3.509×10^{-2}	1.749×10^{-3}
(50, 0, 200, 0, 100, 100, 50, 0)	$\beta_3 = 3.0$	2.000×10^{-2}	8.372×10^{-3}	0
	$\beta_0 = 1.0$	1.224×10^{-1}	1.660×10^{-2}	8.930×10^{-3}
	$\beta_1 = 1.0$	2.895×10^{-1}	3.001×10^{-2}	2.380×10^{-2}
	$\beta_2 = 1.5$	1.000×10^{-1}	3.509×10^{-2}	1.749×10^{-3}
(50, 0, 200, 0, 100, 100, 50, 0)	$\beta_3 = 3.0$	2.000×10^{-2}	8.372×10^{-3}	0
	$\beta_0 = 1.0$	1.224×10^{-1}	1.660×10^{-2}	8.921×10^{-3}
	$\beta_1 = 1.0$	2.895×10^{-1}	3.001×10^{-2}	2.378×10^{-2}
	$\beta_2 = 2.0$	1.000×10^{-1}	3.509×10^{-2}	1.749×10^{-3}
(50, 0, 200, 0, 100, 100, 20, 30)	$\beta_3 = 3.0$	2.000×10^{-2}	8.372×10^{-3}	0
	$\beta_0 = 1.0$	1.224×10^{-1}	7.183×10^{-3}	2.314×10^{-3}
	$\beta_1 = 1.0$	2.895×10^{-1}	1.057×10^{-2}	4.814×10^{-2}
	$\beta_2 = 2.0$	1.000×10^{-1}	1.439×10^{-2}	8.327×10^{-3}
(50, 0, 0, 200, 100, 100, 50, 0)	$\beta_3 = 3.0$	2.000×10^{-2}	5.173×10^{-3}	1.090×10^{-3}
	$\beta_0 = 1.0$	1.224×10^{-1}	1.116×10^{-2}	3.274×10^{-3}
	$\beta_1 = 1.0$	2.895×10^{-1}	1.476×10^{-2}	9.621×10^{-2}
	$\beta_2 = 2.0$	1.000×10^{-1}	2.191×10^{-2}	1.112×10^{-3}
(50, 0, 0, 200, 100, 100, 50, 0)	$\beta_3 = 2.0$	2.000×10^{-2}	7.754×10^{-3}	1.264×10^{-3}
	$\beta_0 = 1.0$	1.224×10^{-1}	7.183×10^{-3}	2.206×10^{-3}
	$\beta_1 = 1.0$	2.895×10^{-1}	1.057×10^{-2}	4.519×10^{-3}
	$\beta_2 = 2.0$	1.000×10^{-1}	1.439×10^{-2}	8.327×10^{-3}
(50, 0, 0, 200, 100, 100, 20, 30)	$\beta_3 = 3.0$	2.000×10^{-2}	5.173×10^{-3}	1.090×10^{-3}

Table 4

Variance Comparison with MLE's Plugged in for Linear Regression

$(n_1, n_{22}, n_{23}n_3, n_{42}, n_{43}, n_5)$	Para	CC				CR				AC			
		MLE	$Sim{:}V_{cc}$	V_{cc}	95% CP	MLE	$Sim{:}V_{cc}{-}Sim{:}V_{cr}$	$V_{cc}{-}V_{cr}$	95% CP	MLE	$Sim{:}V_{cr}{-}Sim{:}V_{ac}$	$V_{cr}{-}V_{ac}$	95% CP
(50, 200, 0, 0, 100, 100, 50, 0)	$\beta_0 = 1.0$	0.988	0.130	0.124	97.2	0.994	6.922×10^{-2}	6.214×10^{-2}	97.4	0.992	1.448×10^{-2}	1.490×10^{-2}	96.8
	$\beta_1 = 1.0$	1.011	0.305	0.292	97.2	0.995	1.500×10^{-1}	1.442×10^{-1}	97.4	0.994	4.252×10^{-2}	3.566×10^{-2}	97.5
	$\beta_2 = 2.0$	2.003	0.112	0.100	97.0	1.987	7.352×10^{-2}	6.392×10^{-2}	96.7	1.981	1.383×10^{-3}	1.101×10^{-3}	96.6
	$\beta_3 = 3.0$	3.000	0.022	0.020	97.1	3.006	1.483×10^{-2}	1.332×10^{-2}	97.7	3.009	$< 1.0 \times 10^{-4}$	$< 1.0 \times 10^{-4}$	98.0
(50, 200, 0, 0, 100, 100, 50, 0)	$\beta_0 = 1.0$	0.988	0.130	0.124	97.2	0.994	6.922×10^{-2}	6.214×10^{-2}	97.4	0.992	1.448×10^{-2}	1.490×10^{-2}	96.8
	$\beta_1 = 1.0$	1.011	0.305	0.292	97.2	0.995	1.500×10^{-1}	1.442×10^{-1}	97.4	0.994	4.252×10^{-2}	3.566×10^{-2}	97.5
	$\beta_2 = 2.0$	2.003	0.112	0.100	97.0	1.987	7.352×10^{-2}	6.392×10^{-2}	96.7	1.981	1.383×10^{-3}	1.101×10^{-3}	96.6
	$\beta_3 = 3.0$	2.500	0.022	0.020	97.1	2.506	1.483×10^{-2}	1.332×10^{-2}	97.7	2.509	$< 1.0 \times 10^{-4}$	$< 1.0 \times 10^{-4}$	98.0
(50, 200, 0, 0, 100, 100, 50, 0)	$\beta_0 = 1.0$	0.988	0.130	0.124	97.2	0.994	6.922×10^{-2}	6.214×10^{-2}	97.4	0.992	1.405×10^{-2}	1.461×10^{-2}	96.8
	$\beta_1 = 1.0$	1.011	0.305	0.292	97.2	0.995	1.500×10^{-1}	1.442×10^{-1}	97.4	0.994	4.117×10^{-2}	3.495×10^{-2}	97.4
	$\beta_2 = 2.0$	2.003	0.112	0.100	97.0	1.987	7.352×10^{-2}	6.392×10^{-2}	96.7	1.981	1.383×10^{-3}	1.100×10^{-3}	96.6
	$\beta_3 = 3.0$	3.000	0.022	0.020	97.1	3.006	1.483×10^{-2}	1.332×10^{-2}	97.7	3.009	$< 1.0 \times 10^{-4}$	$< 1.0 \times 10^{-4}$	98.0
(50, 0, 200, 0, 100, 100, 50, 0)	$\beta_0 = 1.0$	0.988	0.130	0.124	97.3	0.987	1.939×10^{-2}	1.686×10^{-2}	97.4	0.991	8.583×10^{-3}	9.371×10^{-3}	97.2
	$\beta_1 = 1.0$	1.011	0.305	0.292	97.3	1.010	4.199×10^{-2}	3.087×10^{-2}	97.7	1.008	2.295×10^{-2}	2.459×10^{-2}	97.9
	$\beta_2 = 2.0$	2.003	0.112	0.100	97.0	2.003	4.255×10^{-2}	3.509×10^{-2}	97.3	2.005	3.047×10^{-3}	1.784×10^{-3}	97.4
	$\beta_3 = 3.0$	3.000	0.022	0.020	97.1	3.000	1.017×10^{-2}	8.376×10^{-3}	97.1	2.999	$< 1.0 \times 10^{-4}$	$< 1.0 \times 10^{-4}$	97.0
(50, 0, 200, 0, 100, 100, 50, 0)	$\beta_0 = 1.0$	0.988	0.130	0.124	97.3	0.987	1.939×10^{-2}	1.686×10^{-2}	97.4	0.991	8.583×10^{-3}	9.371×10^{-3}	97.2
	$\beta_1 = 1.0$	1.011	0.305	0.292	97.3	1.010	4.199×10^{-2}	3.087×10^{-2}	97.7	1.008	2.295×10^{-2}	2.459×10^{-2}	97.9
	$\beta_2 = 1.5$	1.503	0.112	0.100	97.0	1.503	4.255×10^{-2}	3.509×10^{-2}	97.3	1.503	3.047×10^{-3}	1.784×10^{-3}	97.4
	$\beta_3 = 3.0$	3.000	0.022	0.020	97.1	3.000	1.017×10^{-2}	8.376×10^{-3}	97.1	2.999	$< 1.0 \times 10^{-4}$	$< 1.0 \times 10^{-4}$	97.0
(50, 0, 200, 0, 100, 100, 20, 30)	$\beta_0 = 1.0$	0.988	0.130	0.124	97.3	0.987	1.939×10^{-2}	1.686×10^{-2}	97.4	0.991	8.625×10^{-3}	9.341×10^{-3}	97.2
	$\beta_1 = 1.0$	1.011	0.305	0.292	97.3	1.010	4.199×10^{-2}	3.087×10^{-2}	97.7	1.008	2.305×10^{-2}	2.456×10^{-2}	97.9
	$\beta_2 = 2.0$	2.003	0.112	0.100	97.0	2.003	4.255×10^{-2}	3.509×10^{-2}	97.3	2.005	3.046×10^{-3}	1.784×10^{-3}	97.4
	$\beta_3 = 3.0$	3.000	0.022	0.020	97.1	3.000	1.017×10^{-2}	8.376×10^{-3}	97.1	2.999	$< 1.0 \times 10^{-4}$	$< 1.0 \times 10^{-4}$	97.0
(50, 0, 200, 0, 100, …)	$\beta_0 = 1.0$	0.988	0.130	0.124	97.2	0.983	2.849×10^{-3}	6.160×10^{-3}	96.9	0.987	3.209×10^{-3}	3.779×10^{-3}	96.9
	$\beta_1 = 1.0$	1.011	0.305	0.292	97.2	1.005	5.615×10^{-3}	8.373×10^{-3}	97.4	1.008	8.295×10^{-3}	8.444×10^{-3}	97.5
	$\beta_2 = 2.0$	2.003	0.112	0.100	97.0	1.998	2.241×10^{-4}	1.448×10^{-2}	95.9	2.001	1.159×10^{-4}	8.268×10^{-3}	95.1
	$\beta_0 = 3.0$	0.900	0.030	0.024	97.1	0.083	2.849×10^{-3}	5.160×10^{-3}	94.9	0.007	3.209×10^{-4}	3.089×10^{-3}	94.9

$(n_1, n_{22}, n_{23}n_3, n_{42}, n_{43}, n_5)$	Para	CC				CR				AC			
		MLE	$Sim{:}V_{cc}$	V_{cc}	95% CP	MLE	$Sim{:}V_{cc}{-}Sim{:}V_{cr}$	$V_{cc}{-}V_{cr}$	95% CP	MLE	$Sim{:}V_{cr}{-}Sim{:}V_{ac}$	$V_{cr}{-}V_{ac}$	95% CP
(50, 0, 200, 0, 100, 100, 50, 0)	$\beta_1 = 1.0$	1.011	0.305	0.292	97.2	1.005	5.615×10^{-3}	8.373×10^{-3}	97.4	1.008	8.295×10^{-3}	8.444×10^{-3}	97.5
	$\beta_0 = 1.0$	0.989	0.130	0.104	97.0	0.988	5.20×10^{-3}	1.04×10^{-2}	96.9	0.99	5.19×10^{-3}	8.26×10^{-3}	96.8
	$\beta_1 = 1.0$	1.000	0.095	0.090	97.1	1.004	9.89×10^{-3}	5.39×10^{-2}	94.9	1.007	1.36×10^{-4}	1.08×10^{-2}	94.0
	$\beta_2 = 2.0$	2.003	0.112	0.100	97.0	1.996	5.067×10^{-4}	2.205×10^{-2}	95.3	2.000	2.442×10^{-4}	1.101×10^{-2}	93.6
	$\beta_3 = 2.0$	2.000	0.022	0.020	97.1	2.004	2.720×10^{-4}	7.745×10^{-3}	93.1	2.001	1.257×10^{-4}	1.264×10^{-3}	91.9
(50, 0, 200, 0, 100, 100, 50, 0)	$\beta_0 = 1.0$	0.988	0.130	0.124	97.2	0.983	2.849×10^{-3}	6.160×10^{-3}	97.4	0.987	2.939×10^{-3}	3.632×10^{-3}	96.9
	$\beta_1 = 1.0$	1.011	0.305	0.292	97.2	1.005	5.615×10^{-3}	8.373×10^{-3}	97.7	1.008	7.312×10^{-3}	8.101×10^{-3}	97.6
	$\beta_2 = 2.0$	2.003	0.112	0.100	97.0	1.998	2.241×10^{-4}	1.448×10^{-2}	97.3	2.001	9.505×10^{-5}	8.266×10^{-3}	95.1
	$\beta_3 = 3.0$	3.000	0.022	0.020	97.1	3.003	1.849×10^{-4}	5.171×10^{-3}	97.1	3.001	$< 1.0 \times 10^{-4}$	1.088×10^{-3}	93.9

$Sim{:}V_{cc}$, $Sim{:}V_{cr}$ and $Sim{:}V_{ac}$ are the simulated variances of CC, CR and AC methods. V_{cc}, V_{cr} and V_{ac} are the variances of CC, CR and AC methods using the formula derived in Section 4 with MLE's plug-in. 95% CP is the 95% coverage probability.

Table 5

Simulation for Logistic Regression Model

Method		$\beta_0 = -1.8$	$\beta_1 = 1.0$	$\beta_2 = 2.0$	$\alpha_0 = -0.5$	$\alpha_1 = 2.0$
CC	Bias	0.224	0.483	0.059	0.104	0.273
	SE	0.450	0.723	0.408	0.351	0.672
	SSE	0.466	0.720	0.419	0.351	0.680
	CP%	99.5	99.6	97.8	98.8	98.4
	MSE	0.484	1.268	0.355	0.257	1.000
CR	Bias	−0.036	−0.008	0.059	−0.017	0.056
	SE	0.378	0.544	0.437	0.372	0.687
	SSE	0.408	0.527	0.419	0.339	0.650
	CP%	97.9	98.7	97.9	98.7	97.7
	MSE	0.333	0.555	0.355	0.230	0.847
AC	Bias	−0.035	−0.007	0.055	−0.019	0.051
	SE	0.377	0.528	0.417	0.262	0.539
	SSE	0.402	0.516	0.415	0.250	0.533
	CP%	97.7	98.3	97.6	98.5	97.7
	MSE	0.324	0.533	0.347	0.126	0.571
FD	Bias	−0.020	0.006	0.023	−0.011	0.016
	SE	0.245	0.366	0.223	0.186	0.344
	SSE	0.247	0.364	0.225	0.181	0.337
	CP%	97.4	97.5	97.4	97.4	98.0
	MSE	0.123	0.264	0.102	0.065	0.227

Bias=$\widehat{\beta} - \beta_{\text{true}}$, SE is the mean of the standard error calculated by Louis's formula, SSE is the simulated standard error, CP is the coverage probability, $MSE = Bias^2 + SSE^2$ is the mean square error.

NIH-PA Author Manuscript

Table 6

Lung Cancer Data Analysis

Method	Effect	Model M1			Model M2		
		Estimate	SE	P-value	Estimate	SE	P-value
CC	Intercept	79.008	3.536	<0.001	79.008	3.536	<0.001
	trt	−2.366	3.222	0.463	−2.366	3.222	0.463
	gender	−3.033	3.431	0.377	−3.033	3.431	0.377
	hist	3.011	3.694	0.415	3.011	3.694	0.415
	age	3.049	1.587	0.055	3.049	1.587	0.055
	apex	4.825	5.004	0.335	4.825	5.004	0.335
	recov	−0.485	3.421	0.887	−0.485	3.421	0.887
	σ^2_{MLE}	147.896	27.230	<0.001	147.896	27.230	<0.001
CR	Intercept	81.565	2.912	<0.001	81.573	2.910	<0.001
	trt	0.743	2.543	0.770	0.730	2.540	0.774
	gender	−6.048	2.580	0.019	−6.034	2.576	0.019
	hist	2.003	3.076	0.515	2.007	3.076	0.514
	age	3.640	1.244	0.003	3.641	1.244	0.003
	apex	3.473	6.088	0.568	3.433	6.078	0.572
	recov	−4.676	4.319	0.279	−4.700	4.317	0.276
	σ^2_{MLE}	210.197	25.511	<0.001	210.183	25.525	<0.001
AC	Intercept	81.533	2.768	<0.001	81.535	2.762	<0.001
	trt	0.905	2.556	0.723	0.905	2.550	0.723
	gender	−5.778	2.516	0.021	−5.777	2.508	0.021
	hist	2.067	3.064	0.500	2.074	3.057	0.497
	age	3.527	1.234	0.004	3.532	1.232	0.004
	apex	3.227	6.168	0.601	3.199	6.154	0.603
	recov	−5.326	4.362	0.222	−5.343	4.348	0.219
	σ^2_{MLE}	208.812	25.667	<0.001	208.436	25.573	<0.001