Statistical Science
2010, Vol. 25, No. 4, 517–532
DOI: 10.1214/10-STS344

Block-Conditional Missing at Random Models for Missing Data

Yan Zhou, Roderick J. A. Little and John D. Kalbfleisch

Abstract. Two major ideas in the analysis of missing data are (a) the EM algorithm [Dempster, Laird and Rubin, *J. Roy. Statist. Soc. Ser. B* **39** (1977) 1–38] for maximum likelihood (ML) estimation, and (b) the formulation of models for the joint distribution of the data Z and missing data indicators M, and associated "missing at random" (MAR) condition under which a model for M is unnecessary [Rubin, *Biometrika* **63** (1976) 581–592]. Most previous work has treated Z and M as single blocks, yielding selection or pattern-mixture models depending on how their joint distribution is factorized. This paper explores "block-sequential" models that interleave subsets of the variables and their missing data indicators, and then make parameter restrictions based on assumptions in each block. These include models that are not MAR. We examine a subclass of block-sequential models we call block-conditional MAR (BCMAR) models, and an associated block-monotone reduced likelihood strategy that typically yields consistent estimates by selectively discarding some data. Alternatively, full ML estimation can often be achieved via the EM algorithm. We examine in some detail BCMAR models for the case of two multinomially distributed categorical variables, and a two block structure where the first block is categorical and the second block arises from a (possibly multivariate) exponential family distribution.

Key words and phrases: Block-sequential missing data models, block-conditional MAR models, EM algorithm, categorical data.

1. INTRODUCTION

Missing values arise in empirical studies for many reasons, including unavailability of the measurements, respondents refusing to answer certain items on a questionnaire, and attrition in longitudinal studies. Complete case (CC) analysis, which omits information in the cases with missing values, is inefficient and potentially biased, especially if the subjects included in the analysis are systematically different from those excluded in terms of one or more key variables. Approaches that incorporate information in the incomplete cases include nonresponse weighting (Little and Rubin, 2002, Chapter 3); multiple imputation (MI), where missing values are replaced by multiple sets of plausible values (Rubin, 1987; Little and Rubin, 2002, Chapter 5); weighted estimating equation (WEE) methods (Lipsitz, Ibrahim and Zhao, 1999); and methods based on the likelihood for a model for the data, such as maximum likelihood (ML) or fully Bayes modeling. We focus here on the ML approach, although our models could also be analyzed using Bayesian or MI methods.

Rubin's (1976) theory on modeling the missing-data mechanism was a key development in estimation with incomplete data. Rubin (1976) formalized the concept of missing-data mechanisms by treating the missing-data indicators as random variables and assigning them a distribution. Specifically, let $Z = (Z_{ij})$

Yan Zhou is Mathematical Statistician, Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, Maryland, USA (e-mail: yan.zhou@fda.hhs.gov). Roderick J. A. Little is Professor, Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, USA (e-mail: rlittle@umich.edu). John D. Kalbfleisch is Professor, Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, USA (e-mail: jdkalbfl@umich.edu).

denote a rectangular $n \times p$ data set; the ith row is $Z_i = (Z_{i1}, \ldots, Z_{ip})$, where Z_{ij} is the jth observation for subject i. Let $M = (M_{ij})$ be a missing data indicator matrix with the ith row $M_i = (M_{i1}, \ldots, M_{ip})$, such that M_{ij} is 1 if Z_{ij} is missing and M_{ij} is 0 if Z_{ij} is present. We assume that (Z_i, M_i), $i = 1, \ldots, n$, are independent and identically distributed. In Rubin (1976), the joint distribution is factored as

$$f(Z_i, M_i|\theta, \psi) = f(Z_i|\theta) f(M_i|Z_i, \psi), \tag{1.1}$$

where $f(Z_i|\theta)$ represents the model for the data without missing values, $f(M_i|Z_i, \psi)$ models the missing-data mechanism, and (θ, ψ) denotes unknown parameters. When missingness does not depend on the values of the data Z, missing or observed, that is, if

$$f(M_i|Z_i, \psi) = f(M_i|\psi) \quad \text{for all } Z_i, \psi,$$

the data are called missing completely at random (MCAR). With the exception of some planned missing-data designs, MCAR is a strong assumption, and missingness often depends on the observed and/or unobserved data. Let $Z_{\text{obs},i}$ denote the observed component of Z_i and $Z_{\text{mis},i}$ the missing component. A less restrictive assumption is that missingness depends only on the observed values $Z_{\text{obs},i}$, and not on the missing values $Z_{\text{mis},i}$. That is,

$$f(M_i|Z_i, \psi) = f(M_i|Z_{\text{obs},i}, \psi) \quad \text{for all } Z_{\text{mis},i}, \psi.$$

The missing-data mechanism is then called missing at random (MAR). The mechanism is called missing not at random (MNAR) if the distribution of M depends on the missing values in the data matrix Z.

The observed data consist of the values of the variables (Z_{obs}, M) and the distribution of the observed data is obtained by integrating Z_{mis} out of the joint density of $Z = (Z_{\text{obs}}, Z_{\text{mis}})$ and M. That is, for unit i,

$$\begin{aligned} &f(Z_{\text{obs},i}, M_i|\theta, \psi) \\ &\quad = \int f(Z_{\text{obs},i}, Z_{\text{mis},i}|\theta) \\ &\qquad \cdot f(M_i|Z_{\text{obs},i}, Z_{\text{mis},i}, \psi)\, dZ_{\text{mis},i}. \end{aligned} \tag{1.2}$$

The full likelihood of θ and ψ is any function of θ and ψ proportional to the product of (1.2) over observations i:

$$L_{\text{full}}(\theta, \psi|Z_{\text{obs}}, M) \propto \prod_{i=1}^{n} f(Z_{\text{obs},i}, M_i|\theta, \psi).$$

The missing-data mechanism is called ignorable if it is MAR and if in addition, the parameter space for (θ, ψ) is a Cartesian product space $\Theta \times \Psi$ where $\theta \in \Theta$ and $\psi \in \Psi$. Likelihood-based inferences for θ can then be based on

$$L_{\text{ign}}(\theta|Z_{\text{obs}}) \propto \prod_{i=1}^{n} f(Z_{\text{obs},i}|\theta),$$

the ignorable likelihood of θ based on the observed data Z_{obs} (Rubin, 1976). Many methods of handling missing data assume missingness is MCAR or MAR. If this is assumed, the missing-data mechanism can be ignored and we only need to model the observed data Z_{obs} to derive likelihood-based inferences for θ. However, these inferences are subject to bias when the data are not MAR.

Equation (1.1) is sometimes called a selection model factorization of the joint distribution of (Z_i, M_i) because of connections with the econometric literature on selection bias (Heckman, 1976). Clearly other factorizations are possible. In particular, pattern-mixture models (Little, 1993) factor the joint distribution as

$$f(Z_i, M_i|\varphi, \pi) = f(M_i|\pi) f(Z_i|M_i, \varphi), \tag{1.3}$$

which models the distribution of Z_i for each pattern of missing data.

Both selection and pattern-mixture models treat the variables Z_i and missing-data indicators M_i as single blocks. Little attention has been paid to models that disaggregate these blocks based on subsets of variables and their missing-data indicators. One such class of models is generated by writing $Z_i = (Z_{i(1)}, Z_{i(2)}, \ldots, Z_{i(B)})$ where $Z_{i(j)}$ is a subset of the variables, with corresponding missing-data indicators $M_i = (M_{i(1)}, M_{i(2)}, \ldots, M_{i(B)})$. For convenience, define the "history" up to block j for unit i as

$$\mathcal{H}_{i(j)} = \big(Z_{i(1)}, M_{i(1)}, \ldots, Z_{i(j)}, M_{i(j)}\big)$$

and factor the joint distribution as

$$\begin{aligned} &f(Z_i, M_i|\theta, \psi) \\ &\quad = f\big(Z_{i(1)}, M_{i(1)}|\theta^{(1)}, \psi^{(1)}\big) \\ &\qquad \cdot f\big(Z_{i(2)}, M_{i(2)}|\mathcal{H}_{i(1)}, \theta^{(2)}, \psi^{(2)}\big) \\ &\qquad \cdot \cdots \cdot f\big(Z_{i(B)}, M_{i(B)}|\mathcal{H}_{i(B-1)}, \theta^{(B)}, \psi^{(B)}\big). \end{aligned} \tag{1.4}$$

We call models based on the factorization (1.4) *block-sequential missing data models*. The set $(Z_{i(j)}, M_{i(j)})$ in the jth block might be modeled using the selection or pattern-mixture factorization, yielding combinations of (1.1) and (1.3). This approach to modeling might be seen as natural when the blocks unfold sequentially in time, or if they follow a causal sequence, and the variables in a block are conditioned on prior

variables in time or in the causal chain. Along these lines, Robins and Gill (1997) and Robins (1997) argue that MAR is hard to justify causally when data do not have a monotone pattern, and discuss alternative factorizations that have a readier causal interpretation.

Various modeling assumptions might be incorporated in (1.4). In this article we consider a particular form of potentially MNAR models based on (1.4) with specific assumptions concerning the dependence of the distribution of the variables in each block on the history. Specifically, we assume that in the jth block, the joint distribution of $(Z_{i(j)}, M_{i(j)}|\mathcal{H}_{i(j-1)})$ can be factorized as follows (parameters are left implicit):

$$\begin{aligned} &f\big(Z_{i(j)}, M_{i(j)}|\mathcal{H}_{i(j-1)}\big) \\ &\quad = f\big(Z_{i(j)}|\mathcal{H}_{i(j-1)}\big) f\big(M_{i(j)}|\mathcal{H}_{i(j-1)}, Z_{i(j)}\big), \end{aligned} \tag{1.5}$$

where

$$f\big(Z_{i(j)}|\mathcal{H}_{i(j-1)}\big) = f\big(Z_{i(j)}|Z_{i(1)}, \ldots, Z_{i(j-1)}\big),$$

$$f\big(M_{i(j)}|\mathcal{H}_{i(j-1)}, Z_{i(j)}\big) = f\big(M_{i(j)}|\mathcal{H}_{i(j-1)}, Z_{\mathrm{obs},i(j)}\big),$$

and $Z_{\mathrm{obs},i(j)}$ denotes the observed components of $Z_{i(j)}$. That is, the distribution of $Z_{i(j)}$ given the previous variables depends only on the previous Z's, not the previous M's, and the distribution of $M_{i(j)}$ can depend on previous Z's, M's and $Z_{\mathrm{obs},i(j)}$, but not on the missing components of $Z_{i(j)}$, say, $Z_{\mathrm{mis},i(j)}$. We call models of the form (1.5) *block-conditional MAR* (*BCMAR*), since each block would be MAR if values of Z in previous blocks were fully observed.

For $B = 2$ blocks, (1.5) reduces to

$$\begin{aligned} &f(Z_i, M_i|\theta, \psi) \\ &\quad = f\big(Z_{i(1)}|\theta^{(1)}\big) f\big(M_{i(1)}|Z_{\mathrm{obs},i(1)}, \psi^{(1)}\big) \\ &\qquad \cdot f\big(Z_{i(2)}|Z_{i(1)}, \theta^{(2)}\big) \\ &\qquad \cdot f\big(M_{i(2)}|M_{i(1)}, Z_{i(1)}, Z_{\mathrm{obs},i(2)}, \psi^{(2)}\big), \end{aligned} \tag{1.6}$$

where $Z_{i(1)}$ is MAR, ignoring information about $Z_{i(2)}$ and $M_{i(2)}$, and missingness of $Z_{i(2)}$ depends on the observed components of $Z_{i(2)}$, observed and unobserved value of $Z_{i(1)}$ and on $M_{i(1)}$. This mechanism is not in general MAR, since missingness of $Z_{i(2)}$ is allowed to depend on missing values of $Z_{\mathrm{mis},i(1)}$. For the particular case where $Z_{i(1)}$ and $Z_{i(2)}$ are single variables, this reduces to the simpler form

$$\begin{aligned} &f(Z_i, M_i|\theta, \psi) \\ &\quad = f\big(Z_{i(1)}|\theta^{(1)}\big) f\big(M_{i(1)}|\psi^{(1)}\big) \\ &\qquad \cdot f\big(Z_{i(2)}|Z_{i(1)}, \theta^{(2)}\big) \\ &\qquad \cdot f\big(M_{i(2)}|M_{i(1)}, Z_{i(1)}, \psi^{(2)}\big), \end{aligned} \tag{1.7}$$

because of the MAR condition in each block. In this case, $Z_{i(1)}$ is MCAR and, given $Z_{i(1)}, M_{i(1)}, Z_{i(2)}$ is also MAR. In Section 2 we describe inference for BCMAR models based on a block-monotone reduced likelihood, where the conditional distribution of the variables in each block, given the variables in previous blocks, is computed using only the subset of cases for which the variables in previous blocks are fully observed. This reduced likelihood is related but not quite the same as a partial likelihood as defined by Cox (1975). This reduced likelihood does not require a model for the distribution of the missing-data indicators M. This is a useful property, since specifying models for M can be challenging, and results are vulnerable to misspecification. The block-monotone reduced likelihood becomes the full likelihood when data have a particular pattern, which we call *block monotone*.

Use of the block-monotone reduced likelihood generally involves a loss of information, and an interesting question is how much information is lost; the remainder of the paper examines this question in the context of simple bivariate examples. We analyze in detail the model (1.7) for case of bivariate categorical Z, where the complete cases form a 2-way contingency table, and the incomplete cases form supplemental margins (see, for example, Little and Rubin, 2002, Chapter 13). In addition, we give a less detailed analysis of a more general example with two blocks where the distribution of $Z_{i(2)}$ is from the exponential family.

The EM algorithm (Dempster, Laird and Rubin, 1977), a ubiquitous algorithm for ML estimation from incomplete data and the topic of this special issue, plays a useful role in fitting these models. EM is particularly appealing for categorical data, since the Poisson and multinomial distributions for modeling count data yield complete data loglikelihoods that are linear in the cell counts. Consequently, the E step of EM consists of replacing the complete-data cell counts by conditional expectations given the observed data, in effect distributing the supplemental margins into the full table according to current estimates of the cell probabilities. The M step of EM is the same as complete-data ML estimation based on the data filled in by the E step. This approach to estimation for count data with some grouped counts was first established as ML by Hartley (1958). The application to a (2×2) table with supplemental margins was considered by Chen and Fienberg (1974), and extended to the general class of loglinear models by Fuchs (1982).

For some hierarchical loglinear models the M step of EM requires iteration, so EM involves double iteration.

The usual approach is the Deming–Stephan algorithm, also known as iterative proportional fitting (Bishop, Fienberg and Holland, 1975). If the M step is restricted to just one iteration of Deming–Stephan, the result is an example of an ECM (Expectation Conditional Maximization) algorithm, which achieves similar theoretical properties to EM with just a single iterative loop (Meng and Rubin, 1993; Little and Rubin, 2002). EM is also useful for fitting MNAR models for contingency tables (Baker and Laird, 1985; Fay, 1986; Rubin, Stern and Vehovar, 1995; Little and Rubin 2002, Section 15.7). As shown below, EM also plays a useful role for BCMAR models.

In Section 3, we consider ML estimation for a BCMAR model for bivariate categorical data, where $Z = (Z_{(1)}, Z_{(2)})$ are assumed to have a multinomial distribution. The results are surprising. The block-monotone reduced ML estimates of the parameters of the joint distribution of $(Z_{(1)}, Z_{(2)})$ (as discussed in Section 2) are computed noniteratively from the monotone pattern, excluding the data with $Z_{(2)}$ observed and $Z_{(1)}$ missing. These are in fact the full ML estimates, providing corresponding estimates of the parameters of the missing-data mechanism all lie in the admissible range [0, 1]. If not, then the data with $Z_{(2)}$ observed and $Z_{(1)}$ missing enter into the full ML estimates, and an iterative algorithm such as EM is needed to compute them. In Section 4, a restricted version of the BCMAR model is introduced where missingness of $Z_{(2)}$ depends on the perhaps unobserved value of $Z_{(1)}$ but not on whether $Z_{(1)}$ is missing. Some numerical examples are presented in Section 5 to compare unrestricted and restricted BCMAR models and MAR models and to illustrate when the block-monotone reduced ML estimates in the BCMAR models are full ML. A real data example is given in Section 6. Section 7 explores a more general example of a BCMAR model with two blocks, in which the possibly vector valued variable $Z_{(2)}$ arises from a distribution in the exponential family. Section 8 reviews the ideas of the article and outlines extensions to other missing-data problems.

2. ESTIMATION OF BLOCK-CONDITIONAL MAR MODELS USING A REDUCED LIKELIHOOD

For any BCMAR model, define the *block-monotone reduced likelihood* to be

$$L_{\mathrm{bm}}(\theta) = \prod_{j=1}^{B} \prod_{i \in Q_j} f\big(Z_{\mathrm{obs},i(j)} | Z_{i(1)}, Z_{i(2)}, \ldots, Z_{i(j-1)}, \theta^{(j)}\big), \tag{2.1}$$

where Q_j is the subset of cases with $Z_{i(1)}, Z_{i(2)}, \ldots, Z_{i(j-1)}$ fully observed, that is, $M_{i(1)} = M_{i(2)} = \cdots = M_{i(j-1)} = 0$. Under usual regularity conditions, the estimator of θ that maximizes $L_{\mathrm{bm}}(\theta)$ has the same properties as maximum likelihood, in that it is consistent and asymptotically normal with an asymptotic covariance matrix estimated by $I(\hat{\theta})^{-1}$ where $I(\theta) = -\partial^2 \log L_{\mathrm{bm}}(\theta)/\partial\theta^T\,\partial\theta$. These results can be obtained using conditional arguments similar to those of Cox (1975) in his examination of partial likelihood.

We prove this property for the special case of $B = 2$ blocks; the extension to more than two blocks is straightforward. The observed-data likelihood for the two blocks can be written

$$\begin{aligned} L_{\mathrm{obs}}(\theta, \psi) &= \prod_{i=1}^{n} \{ f\big(Z_{\mathrm{obs},i(1)}, M_{i(1)} | \theta, \psi\big) \\ &\quad \cdot \big[f\big(Z_{\mathrm{obs},i(2)}, M_{i(2)} | Z_{\mathrm{obs},i(1)}, M_{i(1)} = 0, \theta, \psi\big) \big]^{\delta_i} \\ &\quad \cdot \big[f\big(Z_{\mathrm{obs},i(2)}, M_{i(2)} | Z_{\mathrm{obs},i(1)}, M_{i(1)}, \theta, \psi\big) \big]^{1-\delta_i} \}, \end{aligned} \tag{2.2}$$

where $\delta_i = I(M_{i(1)} = 0)$. Note that the second term in the product refers to the cases for which $i \in Q_2$. Consider the pseudo-likelihood generated by the first two terms in the product (2.2). Let $\gamma = (\theta, \psi)$, and denote the corresponding scores as

$$S_{i(1)} = \frac{\partial}{\partial \gamma} \log f\big(Z_{\mathrm{obs},i(1)}, M_{i(1)} | \theta, \psi\big)$$

and

$$\begin{aligned} S_{i(2)} &= \delta_i \frac{\partial}{\partial \gamma} \\ &\quad \cdot \log f\big(Z_{\mathrm{obs},i(2)}, M_{i(2)} | Z_{\mathrm{obs},i(1)}, M_{i(1)} = 0, \theta, \psi\big). \end{aligned}$$

Under usual regularity conditions for the appropriate conditional densities, it is now easily seen that $E[S_{i(j)}] = 0$ and $E[S_{i(j)}^2] = -E[\partial S_{i(j)}/\partial\gamma]$ where $j = 1, 2$. Finally, by conditioning on $Z_{\mathrm{obs},i(1)}, M_{i(1)}$, it can be seen that $E[S_{i(1)} S_{i(2)}] = 0$ so that the scores are uncorrelated. It follows that

$$\sum_{i=1}^{n} \big[S_{i(1)}(\theta, \psi) + S_{i(2)}(\theta, \psi) \big] = 0 \tag{2.3}$$

is an unbiased estimating equation with asymptotic properties similar to those of a likelihood score equation. Under i.i.d. assumptions for the data $\{(Z_{i(1)},$

$M_{i(1)}, Z_{i(2)}, M_{i(2)}), i = 1, \ldots, n\}$, the central limit theorem applies to the total score and a Taylor expansion gives the usual asymptotic normal results for the estimators $\hat{\theta}, \hat{\psi}$ that arise as a solution to (2.3). Further, the asymptotic variance of $\hat{\theta}, \hat{\psi}$ can be estimated as the inverse of the usual observed information. Finally, we note that

$$\begin{aligned} L_{\text{obs}}(\theta, \psi) &= \prod_{i=1}^{n} f\big(Z_{\text{obs},i(1)}|\theta^{(1)}\big) f\big(M_{i(1)}|Z_{\text{obs},i(1)}, \psi^{(1)}\big) \\ &\quad \cdot \prod_{i \in Q_2} f\big(Z_{\text{obs},i(2)}|Z_{i(1)}, \theta^{(2)}\big) \\ &\qquad \cdot f\big(M_{i(2)}|Z_{i(1)}, M_{i(1)} = 0, Z_{\text{obs},i(2)}, \psi^{(2)}\big) \\ &\quad \cdot \prod_{i \notin Q_2} f\big(Z_{\text{obs},i(2)}, M_{i(2)}|Z_{\text{obs},i(1)}, M_{i(1)}, \theta, \psi\big), \end{aligned}$$

where the factorization of the first two products into distinct components for θ and ψ is a result of the BCMAR assumptions. Rearranging terms, we can write

$$L_{\text{obs}}(\theta, \psi) = L_{\text{bm}}(\theta) \times L_{\text{M}}(\psi) \times L_{\text{rest}}(\theta, \psi),$$

where

$$\begin{aligned} L_{\text{bm}}(\theta) &= \prod_{i=1}^{n} f\big(Z_{\text{obs},i(1)}|\theta^{(1)}\big), \\ &\quad \cdot \prod_{i \in Q_2} f\big(Z_{\text{obs},i(2)}|Z_{i(1)}, \theta^{(2)}\big) \\ L_{\text{M}}(\psi) &= \prod_{i=1}^{n} f\big(M_{i(1)}|Z_{\text{obs},i(1)}, \psi^{(1)}\big) \\ &\quad \cdot \prod_{i \in Q_2} f\big(M_{i(2)}|Z_{i(1)}, M_{i(1)} = 0, \\ &\qquad Z_{\text{obs},i(2)}, \psi^{(2)}\big), \\ L_{\text{rest}}(\theta, \psi) &= \prod_{i \notin Q_2} f\big(Z_{\text{obs},i(2)}, M_{i(2)}|Z_{\text{obs},i(1)}, \\ &\qquad M_{i(1)}, \theta, \psi\big). \end{aligned}$$

It can then be easily seen that the observed information matrix based on the first two components is diagonal in the parameters, and the asymptotic results for θ can be determined from $L_{\text{bm}}(\theta)$ as described above.

The block-monotone reduced likelihood inference drops the components $L_{\text{M}}(\psi)$ and $L_{\text{rest}}(\theta, \psi)$ from the likelihood, and bases inference about θ on the remaining term $L_{\text{bm}}(\theta)$. This provides a convenient approach to inference, since the block-monotone reduced likelihood does not involve the distributions of the missing-data indicators, and, hence, these distributions do not need to be specified. Correctly specifying these distributions is not easy, and estimates of θ are vulnerable to their misspecification.

We say that $Z_i = (Z_{i(1)}, Z_{i(2)}, \ldots, Z_{i(B)})$ have a *block monotone* pattern if, for all j, $Z_{i(j-1)}$ is fully observed whenever $Z_{i(j)}$ has at least one observed component. Note that block monotonicity is weaker than a monotone pattern for all the variables, since the variables within each block do not necessarily have a monotone pattern. If the data have a block monotone pattern, the term $L_{\text{rest}}(\theta, \psi)$ is no longer present, and the block-monotone reduced likelihood is equivalent to the full likelihood for inference about θ, providing the parameters θ and ψ are distinct. In other situations, dropping the term $L_{\text{rest}}(\theta, \psi)$ involves a loss of information, so the estimates are not in general fully efficient compared with full ML. We explore this potential loss in efficiency for some simple models in the remainder of this article.

3. UNRESTRICTED BCMAR MODELS FOR BIVARIATE CATEGORICAL DATA

We consider data with $B = 2$, $Z = (Z_{(1)}, Z_{(2)})$ where $Z_{(1)}$ and $Z_{(2)}$ are categorical variables with J and K categories respectively. Both $Z_{(1)}$ and $Z_{(2)}$ may be missing, so there are four missing-data patterns. Let $r = 0, 1, 2, 3$ index the missing-data patterns and let P_r denote the set of sample cases with pattern type $r, r = 0, \ldots, 3$ (see Table 1). Let n_r denote the number of cases in the sample with pattern r and $n = \sum_r n_r$ denote the total sample size.

For categorical $Z_{(1)}$ and $Z_{(2)}$ with J and K levels, data in P_0 can be arranged as a $J \times K$ contingency table, and the data in P_1 and P_2 form supplemental $J \times 1$ and $1 \times K$ margins. Let $n_{(0),jk}$ be the count of complete cases with $Z_{(1)} = j, Z_{(2)} = k$, $n_{(1),j+}$ be the count of cases with $Z_{(1)} = j$ and $Z_{(2)}$ missing, $n_{(2),+k}$ be the count of cases with $Z_{(2)} = k$ and $Z_{(1)}$ missing, and $n_{(3),++}$ be the count of cases with both $Z_{(1)}$ and $Z_{(2)}$ missing. The data are displayed in Table 2.

TABLE 1
Missing-data pattern for two variables

Pattern		
P_0	■	■
P_1	■	?
P_2	?	■
P_3	?	?

TABLE 2
Notation for a $J \times K$ table with supplemental margins for both variables

		$Z_{(2)}$					
		1	**2**	**…**	**…**	***K***	**Missing**
$Z_{(1)}$	**1**	$n_{(0),11}$	$n_{(0),12}$	…	…	$n_{(0),1K}$	$n_{(1),1+}$
	2	$n_{(0),21}$	$n_{(0),22}$	…	…	$n_{(0),2K}$	$n_{(1),2+}$
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	J	$n_{(0),J1}$	$n_{(0),J2}$	…	…	$n_{(0),JK}$	$n_{(1),J+}$
	Missing	$n_{(2),+1}$	$n_{(2),+2}$	…	…	$n_{(2),+K}$	$n_{(3),++}$

Note that $n_0 = \sum_{j=1}^{J}\sum_{k=1}^{K} n_{(0),jk}$, $n_1 = \sum_{j=1}^{J} n_{(1),j+}$, $n_2 = \sum_{k=1}^{K} n_{(2),+k}$, and $n_3 = n_{(3),++}$.

The parameters of interest are $\theta = \{\theta_{jk}\}$, where $\theta_{jk} = P(Z_{(1)} = j, Z_{(2)} = k)$ with $\sum_{j=1}^{J}\sum_{k=1}^{K}\theta_{jk} = 1$. The MAR assumption for these data implies that

$$
\begin{aligned}
&P\big(M_{(1)} = M_{(2)} = 1 | Z_{(1)} = j, Z_{(2)} = k\big) = \upsilon,\\
&P\big(M_{(1)} = 0, M_{(2)} = 1 | Z_{(1)} = j, Z_{(2)} = k\big) = \upsilon_j^{(0)},\\
&P\big(M_{(1)} = 1, M_{(2)} = 0 | Z_{(1)} = j, Z_{(2)} = k\big) = \upsilon_k^{(1)},\\
&P\big(M_{(1)} = M_{(2)} = 0 | Z_{(1)} = j, Z_{(2)} = k\big)\\
&\quad = 1 - \upsilon - \upsilon_j^{(0)} - \upsilon_k^{(1)},
\end{aligned}
$$

where $1 \le j \le J, 1 \le k \le K$ and $M_{(1)}$ and $M_{(2)}$ are missing-data indicators for $Z_{(1)}$ and $Z_{(2)}$ with 1 and 0 denoting missing and observed values respectively (see Little and Rubin, 2002, Example 1.19). In this case, $\zeta = \{\upsilon, \upsilon_j^{(0)}, \upsilon_k^{(1)}\}$ represent nuisance parameters for the missing-data mechanism. Under MAR, the likelihood factors into distinct components of θ and ζ; ML estimation of θ under MAR involves all the observed data and typically requires an iterative algorithm such as EM (Little and Rubin, 2002, Chapter 13).

We consider as an alternative to MAR the following BCMAR model (1.7), which incorporates the assumption that $Z_{(1)}$ is MCAR and missingness of $Z_{(2)}$ depends on $Z_{(1)}$ and $M_{(1)}$:

$$
\begin{aligned}
&P\big(M_{(1)} = 1 | Z_{(1)} = j, Z_{(2)} = k\big) = \phi,\\
&P\big(M_{(2)} = 1 | M_{(1)} = 0, Z_{(1)} = j, Z_{(2)} = k\big)\\
&\quad = \phi_j^{(0)},\\
&P\big(M_{(2)} = 1 | M_{(1)} = 1, Z_{(1)} = j, Z_{(2)} = k\big)\\
&\quad = \phi_j^{(1)},
\end{aligned}
\tag{3.1}
$$

where $1 \le j \le J, 1 \le k \le K$. Here $\Phi = \{\phi, \phi_j^{(0)}, \phi_j^{(1)}\}$ are nuisance parameters corresponding to the missing-data mechanism. The number of parameters in this model is $JK + 2J$, whereas the degrees of freedom of the data are $JK + J + K$, which comprise JK for the complete cases, plus J for the supplemental margin on $Z_{(1)}$, plus K for the supplemental margin on $Z_{(2)}$, plus 1 for the number of cases with $Z_{(1)}$ and $Z_{(2)}$ both missing, minus 1 for the total which is considered fixed at n. When $J = K$, the model has the same number of parameters as degrees of freedom in the data; otherwise, the model has more parameters for $J > K$ or fewer for $J < K$.

Note that if $\phi_j^{(1)} = \phi^{(1)}$ does not depend on j, this reduces to a restricted MAR model in which $Z_{(1)}$ is MCAR and missingness of $Z_{(2)}$ depends on $M_{(1)}$, and only depends on $Z_{(1)}$ for the pattern with $Z_{(1)}$ observed. A likelihood ratio test could be used to test this restricted MAR assumption against the more general BCMAR model and the EM algorithm can be applied to compute the ML estimates (Little and Rubin, 2002, Chapter 13). This restricted MAR model is introduced as a testable submodel of the unrestricted BCMAR model, but we do not view it as particularly appealing substantively, since if missingness of $Z_{(2)}$ depends on $Z_{(1)}$ for the cases with $Z_{(1)}$ observed, one might also expect it to depend on $Z_{(1)}$ for the cases with $Z_{(1)}$ missing. Another submodel of the unrestricted BCMAR model is discussed in Section 4.

3.1 EM Algorithm

The full likelihood for the above model is

$$
\begin{aligned}
&L\big(\theta, \Phi | Z_{\mathrm{obs},(1)}, Z_{\mathrm{obs},(2)}, M\big)\\
&\quad = \prod_{i \in P_0} p(Z_{i(1)}, Z_{i(2)} | \theta)(1 - \phi)\\
&\qquad\qquad \cdot p\big(M_{i(2)} = 0 | Z_{i(1)}, M_{i(1)} = 0, \Phi\big)\\
&\qquad \cdot \prod_{i \in P_1} p(Z_{i(1)} | \theta)(1 - \phi)\\
&\qquad\qquad \cdot p\big(M_{i(2)} = 1 | Z_{i(1)}, M_{i(1)} = 0, \Phi\big)\\
&\qquad \cdot \prod_{i \in P_2} \sum_{Z_{i(1)}} p(Z_{i(1)}, Z_{i(2)} | \theta)\phi\\
&\qquad\qquad\qquad \cdot p\big(M_{i(2)} = 0 | Z_{i(1)}, M_{i(1)} = 1, \Phi\big)\\
&\qquad \cdot \prod_{i \in P_3} \sum_{Z_{i(1)}} p(Z_{i(1)} | \theta)\phi\\
&\qquad\qquad\qquad \cdot p\big(M_{i(2)} = 1 | Z_{i(1)}, M_{i(1)} = 1, \Phi\big).
\end{aligned}
\tag{3.2}
$$

The block-monotone reduced likelihood is

$$L_{\mathrm{bm}}(\theta|Z_{\mathrm{obs},(1)}, Z_{\mathrm{obs},(2)}) = \prod_{i\in P_0} p(Z_{i(1)}, Z_{i(2)}|\theta) \prod_{i\in P_1} p(Z_{i(1)}|\theta), \tag{3.3}$$

which does not model the missing data mechanism, and drops the data for patterns P_2 and P_3. We first consider ML estimation for the full likelihood (3.2), and then discuss the relationship between these ML estimates and the estimates that maximize the block-monotone reduced likelihood (3.3).

One approach to ML estimation is to apply the EM algorithm. To define the E step of EM, let $(\theta_{jk}^{(t)}, \phi_j^{(1)(t)})$ denote the parameter estimates at iteration t, and $n_{(r),jk}^{(t)}$ be the estimate of cell frequency for $Z_{i(1)} = j, Z_{i(2)} = k$ in pattern P_r. The E step distributes the partially classified observations into the table according to the corresponding probabilities:

$$n_{(1),jk}^{(t)} = n_{(1),j+} \cdot \frac{\theta_{jk}^{(t)}}{\theta_{j+}^{(t)}},$$

$$n_{(2),jk}^{(t)} = n_{(2),+k} \cdot \frac{(1-\phi_j^{(1)(t)})\theta_{jk}^{(t)}}{\sum_{j=1}^{J}(1-\phi_j^{(1)(t)})\theta_{jk}^{(t)}},$$

$$n_{(3),jk}^{(t)} = n_{(3),++} \cdot \frac{\phi_j^{(1)(t)}\theta_{jk}^{(t)}}{\sum_{j=1}^{J}\phi_j^{(1)(t)}\theta_{j+}^{(t)}}.$$

The M step calculates new parameters as follows:

$$\theta_{jk}^{(t+1)} = \frac{n_{(0),jk} + n_{(1),jk}^{(t)} + n_{(2),jk}^{(t)} + n_{(3),jk}^{(t)}}{n},$$

$$\phi = \frac{\sum_{i=1}^{n} I(M_{i(1)} = 1)}{n} = \frac{n_2 + n_3}{n},$$

$$\phi_j^{(0)} = \frac{\sum_{i=1}^{n} I(M_{i(1)} = 0, M_{i(2)} = 1, Z_{i(1)} = j)}{\sum_{i=1}^{n} I(M_{i(1)} = 0, Z_{i(1)} = j)} = \frac{n_{(1),j+}}{n_{(1),j+} + n_{(0),j+}},$$

$$\phi_j^{(1)(t+1)} = \frac{\sum_k n_{(3),jk}^{(t)}}{\sum_k n_{(2),jk}^{(t)} + \sum_k n_{(3),jk}^{(t)}}.$$

The E step and M step alternate until the parameter estimates converge.

Note that ϕ and $\{\phi_j^{(0)}\}$ are estimated directly and are unchanged throughout the EM algorithm. Complete-case estimates or estimates arising from the monotone pattern P_0 and P_1 can be chosen as the starting values of $\{\theta_{jk}\}$, and the estimates of $\{\phi_j^{(0)}\}$ or any constant in $(0, 1)$ can be taken as initial values of $\{\phi_j^{(1)}\}$. When $J > K$, the model has more parameters than degrees of the freedom. In this case, multiple maxima may exist, and depending on starting values, the EM algorithm can converge to different estimates. This case will be discussed further below.

3.2 Noniterative ML Estimates

When $J \geq K$, noniterative estimates of the parameters can sometimes be obtained using the factored likelihood method (Little and Rubin, 2002, Chapter 7). We transform the parameters $(\theta_{jk}, \phi, \phi_j^{(0)}, \phi_j^{(1)})$ to

$$\begin{aligned}
\alpha_{(0),jk} &= P(Z_{(1)} = j, Z_{(2)} = k|M_{(1)} = M_{(2)} = 0),\\
\beta_{(1),j+} &= P(Z_{(1)} = j|M_{(1)} = 0, M_{(2)} = 1),\\
\gamma_{(2),+k} &= P(Z_{(2)} = k|M_{(1)} = 1, M_{(2)} = 0),\\
\pi_0 &= P(M_{(1)} = 0, M_{(2)} = 0),\\
\pi_1 &= P(M_{(1)} = 0, M_{(2)} = 1),\\
\pi_2 &= P(M_{(1)} = 1, M_{(2)} = 0),\\
\pi_3 &= P(M_{(1)} = 1, M_{(2)} = 1),
\end{aligned} \tag{3.4}$$

where $1 \leq j \leq J, 1 \leq k \leq K$ and the following constraints apply:

$$\sum_{j=1}^{J}\sum_{k=1}^{K}\alpha_{(0),jk} = 1, \qquad \sum_{j=1}^{J}\beta_{(1),j+} = 1,$$

$$\sum_{k=1}^{K}\gamma_{(2),+k} = 1, \qquad \sum_{r=0}^{3}\pi_r = 1.$$

These parameters correspond to a pattern-mixture factorization, as in (1.3). The components of $(\theta, \Phi) = (\theta_{jk}, \phi, \phi_j^{(0)}, \phi_j^{(1)})$ can be expressed in terms of the new parametrization (3.4) as follows:

$$\begin{aligned}
\theta_{jk} &= \left(\frac{\alpha_{(0),jk}}{\alpha_{(0),j+}}\right)\left(\frac{\pi_0\alpha_{(0),j+} + \pi_1\beta_{(1),j+}}{\pi_0 + \pi_1}\right),\\
\phi &= 1 - \pi_0 - \pi_1,\\
\phi_j^{(0)} &= \frac{\pi_1\beta_{(1),j+}}{\pi_0\alpha_{(0),j+} + \pi_1\beta_{(1),j+}},
\end{aligned} \tag{3.5}$$

and $\{\phi_j^{(1)}, j = 1, \ldots, J\}$ is a solution to the K simultaneous equations

$$\sum_{j=1}^{J}(1-\phi_j^{(1)})\theta_{jk} = P(M_{(2)} = 0, Z_{(2)} = k|M_{(1)} = 1) = \frac{\pi_2}{1-\pi_0-\pi_1}\gamma_{(2),+k},$$

where $\alpha_{(0),j+} = \sum_{k=1}^{K} \alpha_{(0),jk}$.

Letting (φ, π) represent the parameters in (3.4), the likelihood can be written as

$$\begin{aligned} &L(\varphi, \pi | Z_{\mathrm{obs},(1)}, Z_{\mathrm{obs},(2)}, M) \\ &= \prod_{i=1}^{n} p(M_{i(1)}, M_{i(2)}) \\ &\quad \cdot \prod_{i \in p_0} p(Z_{i(1)}, Z_{i(2)} | M_{i(1)} = 0, M_{i(2)} = 0) \\ &\quad \cdot \prod_{i \in p_1} p(Z_{i(1)} | M_{i(1)} = 0, M_{i(2)} = 1) \\ &\quad \cdot \prod_{i \in p_2} p(Z_{i(2)} | M_{i(1)} = 1, M_{i(2)} = 0) \\ &= \prod_{r=0}^{3} \pi_r^{n_r} \prod_{j,k=1}^{J,K} \alpha_{(0),jk}^{n_{(0),jk}} \prod_{j=1}^{J} \beta_{(1),j+}^{n_{(1),j+}} \prod_{k=1}^{K} \gamma_{(2),+k}^{n_{(2),+k}}. \end{aligned}$$

Maximizing the four terms in this likelihood yields

$$\hat{\alpha}_{(0),jk} = \frac{n_{(0),jk}}{n_0}, \qquad \hat{\beta}_{(1),j+} = \frac{n_{(1),j+}}{n_1},$$

$$\hat{\gamma}_{(2),+k} = \frac{n_{(2),+k}}{n_2}, \qquad \hat{\pi}_r = \frac{n_r}{n},$$

where $1 \le j \le J$, $1 \le k \le K$ and $0 \le r \le 3$. Estimates of θ_{jk}, ϕ and $\phi_j^{(0)}$ can then be obtained by substituting the above estimates of $(\varphi, \pi) = (\alpha_{(0),jk}, \beta_{(1),j+}, \gamma_{(2),+k}, \pi_r)$ into equation (3.5). This yields

$$\hat{\theta}_{jk} = \left(\frac{n_{(0),jk}}{n_{(0),j+}}\right)\left(\frac{n_{(0),j+} + n_{(1),j+}}{n_0 + n_1}\right), \tag{3.6}$$

$$\hat{\phi} = 1 - \hat{\pi}_0 - \hat{\pi}_1,$$

$$\hat{\phi}_j^{(0)} = \frac{\hat{\pi}_1 \hat{\beta}_{(1),j+}}{\hat{\pi}_0 \hat{\alpha}_{(0),j+} + \hat{\pi}_1 \hat{\beta}_{(1),j+}}. \tag{3.7}$$

Estimates of $\{\phi_j^{(1)}, j = 1, \ldots, J\}$ can be obtained as solutions of the following K simultaneous equations, provided they are in the parameter space:

$$\sum_{j=1}^{J} (1 - \hat{\phi}_j^{(1)}) \hat{\theta}_{jk} = \frac{\hat{\pi}_2}{1 - \hat{\pi}_0 - \hat{\pi}_1} \hat{\gamma}_{(2),+k}. \tag{3.8}$$

This approach yields ML estimates, providing the estimates lie within the parameter space, that is, the probabilities lie between zero and one. The expressions for $\hat{\theta}_{jk}, \hat{\phi}$ and $\hat{\phi}_j^{(0)}$ always yield estimates in [0, 1]. The equations in (3.8), however, may or may not yield solutions for $\{\phi_j^{(1)}\}$ that lie in [0, 1]. If they do, then estimates from this approach are ML estimates and the ML estimates of θ_{jk}, ϕ and $\phi_j^{(0)}$ are unique. If not, this approach fails to yield ML estimates of the parameters of interest. In this case, however, the EM algorithm can still be used, and whether the ML estimate is unique or not depends on the form of the likelihood. If the likelihood is unimodel, the ML estimate is unique. The solution set for (3.8) depends on whether $J = K$ or $J > K$. When $J = K$ there are J equations for J unknowns. Provided the $J \times J$ matrix, $\hat{\Theta} = (\hat{\theta}_{jk})$, is nonsingular, these equations yield a unique solution that may or may not lie in the parameter space. When $J \ge K$ and $\hat{\Theta}$ has rank $K' < J$, the solution set is a linear subspace of dimension $J - K'$. If the solution space intersects the parameter space $[0, 1]^J$, then this approach yields the ML estimates. For example, consider the case where $J = 3$, $K = 2$ and $\hat{\Theta}$ is of full rank K, the solution set to (3.8) is a straight line. When it intersects the unit cube representing the parameter space, this approach yields unique ML estimates of θ_{jk}, ϕ and $\phi_j^{(0)}$, but any point in $[0, 1]^J$ that is in the solution set of (3.8) is a ML estimate for $\{\phi_j^{(1)}\}$. However, when the solution set does not intersect the unit cube, this method fails to yield the ML estimates of the parameters. The EM algorithm can be implemented to find ML estimates, which may or may not be unique. When $J < K$, noniterative ML estimates do not exist and the EM algorithm can be applied to compute ML estimates.

The closed-form estimates (3.6) of θ are simply the product of the estimated conditional probabilities of $Z_{(2)} = k$ given $Z_{(1)} = j$ from the complete cases and the marginal probabilities of $Z_{(1)} = j$ from the cases with $Z_{(1)}$ observed. These estimates maximize the block-monotone reduced likelihood discussed in Section 2, which drops the data for $Z_{(2)}$ from the pattern P_2 with $Z_{(2)}$ observed and $Z_{(1)}$ missing. One would expect the data in P_2 to provide additional information for the marginal distribution of $Z_{(2)}$, but this is only the case if the data in P_2 are inconsistent with the data on $Z_{(2)}$ from P_0 and P_1, in the sense of yielding estimates of $\{\phi_j^{(1)}\}$ from (3.8) that lie outside the interval [0, 1].

4. A RESTRICTED BCMAR MODEL

In the unrestricted BCMAR model (3.1), the missingness of $Z_{(2)}$ is allowed to depend not only on the (perhaps unobserved) value of $Z_{(1)}$ but also on whether $Z_{(1)}$ is missing or not. If, given the value of $Z_{(1)}$, the probability of $Z_{(2)}$ being missing is assumed the same for the cases with $Z_{(1)}$ observed and missing, we then

have the restricted BCMAR model:

$$
\begin{aligned}
P\big(M_{(1)}=1|Z_{(1)}=j,Z_{(2)}=k\big)&=\phi,\\
P\big(M_{(2)}=1|M_{(1)}=l,Z_{(1)}=j,Z_{(2)}=k\big)&=\phi_j,
\end{aligned}
\tag{4.1}
$$

where $l=1,2$ and $1\le j\le J, 1\le k\le K$. The number of the parameters in this model is $JK+J$ which is always less than the degree of freedom $JK+J+K$ in the data. The explicit estimates in (3.6) are no longer ML estimates of $\{\theta_{jk}\}$, and EM is needed to obtain ML estimates of the parameters. In the E step, the partially classified observations are effectively distributed into the table according to the corresponding estimated probabilities:

$$
n^{(t)}_{(1),jk}=n_{(1),j+}\cdot\frac{\theta^{(t)}_{jk}}{\theta^{(t)}_{j+}},
$$

$$
n^{(t)}_{(2),jk}=n_{(2),+k}\cdot\frac{(1-{\phi_j}^{(t)})\theta^{(t)}_{jk}}{\sum_{j=1}^{J}(1-{\phi_j}^{(t)})\theta^{(t)}_{jk}},
$$

$$
n^{(t)}_{(3),jk}=n_{(3),++}\cdot\frac{{\phi_j}^{(t)}\theta^{(t)}_{jk}}{\sum_{j=1}^{J}{\phi_j}^{(t)}\theta^{(t)}_{j+}}.
$$

In the M step, new estimates are calculated as

$$
\theta^{(t+1)}_{jk}=\frac{n_{(0),jk}+n^{(t)}_{(1),jk}+n^{(t)}_{(2),jk}+n^{(t)}_{(3),jk}}{n},
$$

$$
\phi=\frac{n_2+n_3}{n},
$$

$$
{\phi_j}^{(t+1)}=\frac{\sum_k n^{(t)}_{(1),jk}+\sum_k n^{(t)}_{(3),jk}}{n_{(0),j+}+\sum_k n^{(t)}_{(1),jk}+\sum_k n^{(t)}_{(2),jk}+\sum_k n^{(t)}_{(3),jk}}.
$$

The E step and M step alternate until the parameter estimates converge. Since ϕ is estimable directly and is unchanged throughout the EM algorithm, starting values are only needed for $\{\theta_{jk}\}$ and $\{\phi_j\}$. Complete-case estimates or pooled estimates from the monotone pattern P_0 and P_1 can be used as starting values of $\{\theta_{jk}\}$. Estimates of $\{\phi^{(0)}_j\}$ in (3.7) or any constant in (0, 1) can be taken as initial values of $\{\phi_j\}$.

The restricted BCMAR model (4.1) is a submodel of the unrestricted BCMAR model (3.1) obtained by assuming $\phi^{(0)}_j=\phi^{(1)}_j$. The restricted model is plausible when the mechanism of missingness of $Z_{(1)}$ is relatively unrelated to the mechanism of missingness of $Z_{(2)}$, so the probability that one variable is missing is not thought to be related to whether the other variable is missing. The appeal of the restricted model is that it is more parsimonious and will tend to yield more efficient estimates of the parameters of interest. A likelihood ratio test can be applied to test the restricted BCMAR assumption against the more general unrestricted BCMAR model, and one may favor the restricted BCMAR if this test is not rejected.

5. NUMERICAL EXAMPLES

5.1 Examples with $J=K=2$

For data given in the 2×2 Table 3A with supplemental margins, the noniterative estimates of $\{\theta_{jk}\}$ that drop the data in P_2 are ML estimates under the unrestricted BCMAR model. The estimates of $\{\theta_{jk}\}$ are also close to those in the restricted BCMAR and MAR models which involve all the data (Table 4). However, for data in Table 3B, the marginal distribution of $Z_{(2)}$ in P_2 is substantially different from that in the monotone pattern P_0 and P_1. In this case, the unrestricted BCMAR model yields the estimates of $\{\phi^{(1)}_j\}$ from (3.8) that do not lie between 0 and 1. The EM algorithm applied to all the data is needed to obtain the ML estimates, and the estimates of $\{\theta_{jk}\}$ are different from those in the restricted BCMAR and MAR models (Table 5).

TABLE 3
2×2 tables with supplemental margins for both variables

3A

		$Z_{(2)}$		
		1	**2**	**Missing**
$Z_{(1)}$	**1**	50	150	30
	2	75	75	60
	Missing	28	60	50

3B

		$Z_{(2)}$		
		1	**2**	**Missing**
$Z_{(1)}$	**1**	100	50	30
	2	75	75	60
	Missing	28	60	50

TABLE 4
Estimates of parameters for data in Table 3A

	Parameter of interest				Nuisance parameter				
	θ_{11}	θ_{12}	θ_{21}	θ_{22}	ϕ	$\phi_1^{(0)}$	$\phi_2^{(0)}$	$\phi_1^{(1)}$	$\phi_2^{(1)}$
Unrestricted BCMAR noniterative estimate	0.131	0.392	0.239	0.239	0.239	0.130	0.286	0.113	0.636
EM algorithm	0.131	0.392	0.239	0.239	0.239	0.130	0.286	0.113	0.636
Restricted BCMAR					$\phi_j^{(0)} = \phi_j^{(1)}, j = 1, 2$				
					ϕ	ϕ_1		ϕ_2	
EM algorithm	0.126	0.390	0.238	0.246	0.239	0.157		0.333	
Restricted MAR								$\phi_1^{(1)} = \phi_2^{(1)}$	
					ϕ	$\phi_1^{(0)}$	$\phi_2^{(0)}$	$\phi^{(1)}$	
EM algorithm	0.127	0.398	0.232	0.243	0.239	0.130	0.286	0.362	

5.2 Examples with $J = 3$, $K = 2$

Table 6A and B give data for the case $J = 3$, $K = 2$ for which the solution set to (3.8) is a straight line. The parameter space for $\{\phi_j^{(1)}\}$ is a unit cube, as displayed in Figures 1 and 2. For the data in Table 6A, the solution line does not intersect the cube (Figure 1), so ML estimates in the unrestricted BCMAR model are obtained iteratively using all the data (Table 7). For the data in Table 6B, the marginal distribution of $Z_{(2)}$ in P_2 is similar to that in P_0 and P_1 and the solution line intersects the cube (Figure 2), and the noniterative estimates obtained by dropping the data in P_2, displayed in Table 8, are the ML estimates of $\{\theta_{jk}\}$, although there are multiple ML estimates for $\{\phi_j^{(1)}\}$. ML estimates in the restricted BCMAR and MAR models are unique for both data sets in Table 6.

6. MUSCATINE CORONARY RISK FACTOR STUDY

The Muscatine Coronary Risk Factor Study (MCRF) is a longitudinal study of obesity in 4856 school children. Five cohorts (ages 5–7, 7–9, 9–11, 11–13, 13–15) of boys and girls were measured for height and weight in 1977, 1979 and 1981. Children with relative weight greater than 110 percent of the median weight for their age-gender-height group were classified as obese, and at any time point about 20 percent of the children were obese. We are interested in estimating obesity rates over time and evaluating whether or not these rates differ by gender. The study was first presented by Woolson and Clarke (1984), and further analyses can be found in, for example, Baker (1995), Ekholm and Skinner (1998), Lipsitz, Parzen and Molenberghs (1998) and Birmingham and Fitzmaurice (2002).

TABLE 5
Estimates of parameters for data in Table 3B

	Parameters of interest				Nuisance parameter				
	θ_{11}	θ_{12}	θ_{21}	θ_{22}	ϕ	$\phi_1^{(0)}$	$\phi_2^{(0)}$	$\phi_1^{(1)}$	$\phi_2^{(1)}$
Unrestricted BCMAR noniterative estimate	0.308	0.154	0.269	0.269	0.261	0.167	0.286	2.507	−1.476
EM algorithm	0.297	0.153	0.236	0.314	0.261	0.167	0.286	0.867	0
Restricted BCMAR					$\phi_j^{(0)} = \phi_j^{(1)}, j = 1, 2$				
					ϕ	ϕ_1		ϕ_2	
EM algorithm	0.274	0.175	0.242	0.309	0.261	0.197		0.320	
Restricted MAR							$\phi_1^{(1)} = \phi_2^{(1)}$		
					ϕ	$\phi_1^{(0)}$	$\phi_2^{(0)}$	$\phi^{(1)}$	
EM algorithm	0.279	0.174	0.239	0.308	0.261	0.167	0.286	0.362	

TABLE 6
3×2 *tables with supplemental margins for both variables*

6A

		$Z_{(2)}$		
		1	**2**	**Missing**
$Z_{(1)}$	**1**	100	50	30
	2	75	75	60
	3	32	67	20
	Missing	28	60	50

6B

		$Z_{(2)}$		
		1	**2**	**Missing**
$Z_{(1)}$	**1**	50	150	30
	2	75	75	60
	3	32	67	20
	Missing	28	60	50

The analysis is complicated by the study design. Both cross-sectional and longitudinal information about age trends in obesity rates were present in the data. Due to cohort effects, cross-sectional age trends in obesity rates may be different from longitudinal trends. Ekholm and Skinner (1998) found no statistical evidence of cohort effects. Therefore, in our analyses, cohort effects are assumed negligible and data are pooled across five age-group cohorts. In order to simplify the illustration, we only use the data from the surveys of years 1977 and 1981 (Table 9).

The analysis is further complicated by the substantial nonresponse. Only 40 percent of children provided complete records in 1977 and 1981. In addition to the complete records, there are three nonresponse patterns, specifically, two patterns with one missing response and one pattern with two missing responses. Baker (1995) reported two main reasons for nonresponse: (1) no parental consent form was received and (2) the child was not in school on the examination day. For girls, the missingness of obese status in 1981 is found to depend on the missingness in 1977 using a chi-square test (p-value < 0.0001). Furthermore, girls measured and classified as obese in 1977 were more likely to have missing data in 1981 than those classified as nonobese (p-value < 0.0001 based on a chi-square test). The estimates of girls' obesity rates and missing probabilities in the BCMAR model discussed above are presented in Table 10. For the unrestricted BCMAR model, the estimate from (3.8) of $\{\phi_1^{(1)}, \phi_2^{(1)}\}$ is $(0.274, 0.121)$, which is in the parameter space, so closed form estimates of the parameters are available. A bootstrap approach was used to estimate standard errors. If a bootstrap sample leads to the solutions of $\{\phi_j^{(1)}\}$ from (3.8) that lie outside of the parameter space, the EM algorithm is used to obtain the ML estimates. Among the 1000 bootstrap samples, 23.2% of

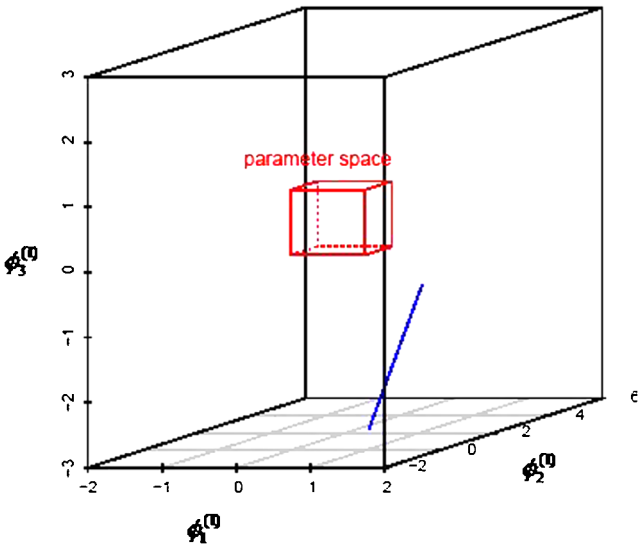

FIG. 1. *Noniterative estimates of* $\phi_j^{(1)}$ *for data in Table 6A.*

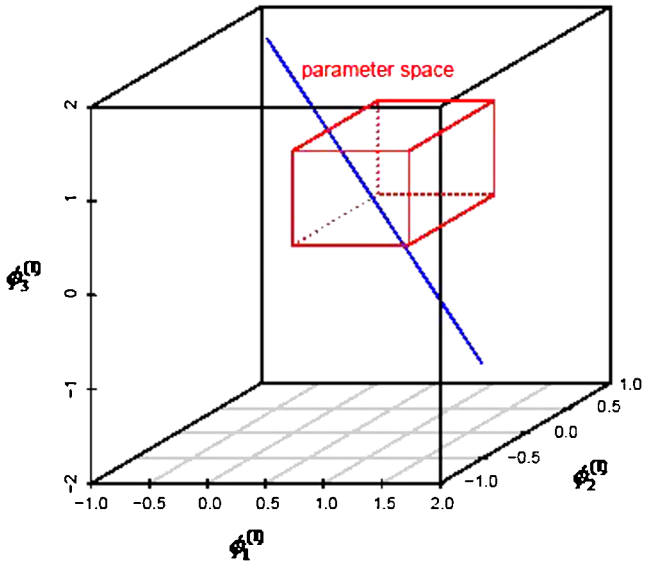

FIG. 2. *Noniterative estimates of* $\phi_j^{(1)}$ *for data in Table 6B.*

TABLE 7
Estimates of parameters for data in Table 6A

	Parameter of interest						Nuisance parameter						
	θ_{11}	θ_{12}	θ_{21}	θ_{22}	θ_{31}	θ_{32}	ϕ	$\phi_1^{(0)}$	$\phi_2^{(0)}$	$\phi_3^{(0)}$	$\phi_1^{(1)}$	$\phi_2^{(1)}$	$\phi_3^{(1)}$
Unrestricted BCMAR													
Noniterative estimate	0.236	0.118	0.206	0.206	0.076	0.158	0.213	0.167	0.286	0.168	*no solution in* $[0,1]^3$		
EM algorithm	0.235	0.117	0.192	0.219	0.071	0.166	0.213	0.167	0.286	0.168	1	0.037	0
Restricted BCMAR							$\phi_j^{(0)} = \phi_j^{(1)}, j = 1,2,3$						
							ϕ	ϕ_1		ϕ_2		ϕ_3	
EM algorithm	0.218	0.126	0.194	0.224	0.069	0.168	0.213	0.196		0.322		0.190	
Restricted MAR											$\phi_1^{(1)} = \phi_2^{(1)} = \phi_3^{(1)}$		
							ϕ	$\phi_1^{(0)}$	$\phi_2^{(0)}$	$\phi_3^{(0)}$	$\phi^{(1)}$		
EM algorithm	0.221	0.127	0.190	0.223	0.070	0.169	0.213	0.167	0.286	0.168	0.362		

the samples yield the solutions of $\{\phi_j^{(1)}\}$ from (3.8) that are outside of the parameter space.

Likelihood ratio tests can be utilized to test the two submodels discussed above against the more general unrestricted BCMAR model. Denote the unrestricted BCMAR model as M1, the restricted BCMAR model as M2 and the restricted MAR model in Section 3 as M3, and let $l_{\max}$ represent the maximized value of the loglikelihood. We find that $-2(l_{\max}(M2) - l_{\max}(M1)) = -2(-4569.823 + 4535.292) = 69.062$, which yields a p-value < 0.0001 when compared to χ_2^2. There is strong evidence that the restricted BCMAR model does not fit the data. On the other hand, $l_{\max}(M3)$ is close to $l_{\max}(M1)$, and we cannot differentiate the restricted MAR model from the unrestricted BCMAR model.

Similarly for the boys, the estimate from (3.8) of $\{\phi_1^{(1)}, \phi_2^{(1)}\}$ in the unrestricted BCMAR model is $(0.228, 0.325)$, which is in the parameter space, and closed form estimates of the parameters are available. Among 1000 bootstrap samples, only 28 samples yield the solutions of $\{\phi_j^{(1)}\}$ from (3.8) outside of the parameter space. The likelihood ratio test yields strong evidence against the restricted BCMAR model, with $-2(l_{\max}(M2) - l_{\max}(M1)) = -2(-4748.48 + 4713.03) = 70.9$ on two degrees of freedom. On the other hand, $l_{\max}(M3)$ is close to $l_{\max}(M1)$, and the restricted MAR model seems to be satisfactory (Table 11).

The models considered above show a small effect on the fitted values of obesity rates and their standard errors. For boys, the marginal distributions of 1981 obe-

TABLE 8
Estimates of parameters for data in Table 6B

	Parameter of interest						Nuisance parameter						
	θ_{11}	θ_{12}	θ_{21}	θ_{22}	θ_{31}	θ_{32}	ϕ	$\phi_1^{(0)}$	$\phi_2^{(0)}$	$\phi_3^{(0)}$	$\phi_1^{(1)}$	$\phi_2^{(1)}$	$\phi_3^{(1)}$
Unrestricted BCMAR													
Noniterative estimate	0.103	0.309	0.188	0.188	0.069	0.144	0.198	0.130	0.286	0.168	*multiple solutions in* $[0,1]^3$		
EM algorithm	0.103	0.309	0.188	0.188	0.069	0.144	0.198	0.130	0.286	0.168	*multiple solutions*		
Restricted BCMAR							$\phi_j^{(0)} = \phi_j^{(1)}, j = 1,2,3$						
							ϕ	ϕ_1		ϕ_2		ϕ_3	
EM algorithm	0.100	0.307	0.189	0.193	0.067	0.144	0.198	0.154		0.328		0.197	
Restricted MAR											$\phi_1^{(1)} = \phi_2^{(1)} = \phi_3^{(1)}$		
							ϕ	$\phi_1^{(0)}$	$\phi_2^{(0)}$	$\phi_3^{(0)}$	$\phi^{(1)}$		
EM algorithm	0.101	0.311	0.184	0.190	0.068	0.146	0.198	0.130	0.286	0.168	0.362		

TABLE 9
Tables of data from muscatine coronary risk factor study

		1981		
		1	**2**	**Missing**
Girls				
	1	701	98	497
1977	**2**	59	111	183
	Missing	408	139	174
Boys				
	1	699	98	566
1977	**2**	72	116	141
	Missing	473	125	196

Notes: 1 = not obese, 2 = obese.

sity rates are quite similar for those with 1977 obesity rates observed or not. If we consider only the cases with 1977 obesity rates observed, the noniterative block-monotone reduced ML estimates of obesity rates for the unrestricted BCMAR model are ML estimates, and these are close to ML estimates in the restricted BCMAR and MAR models. Furthermore, $\hat{\phi}_1^{(0)}$ and $\hat{\phi}_2^{(0)}$ are close to one another, which suggests a MCAR mechanism. As a consequence, complete-case estimates of obesity rates are also similar to those in three models considered above. For girls, for the same reason, noniterative block-monotone reduced ML estimates of obesity rates for the unrestricted BCMAR model are ML estimates and are close to those in the restricted BCMAR and MAR models. However, $\hat{\phi}_1^{(0)}$ and $\hat{\phi}_2^{(0)}$ are quite different, and, as a consequence, complete-case estimates of obesity rates are not similar to those in the other three models.

7. TWO BLOCK BCMAR DATA WITH OUTCOMES FROM THE EXPONENTIAL FAMILY DISTRIBUTION

Suppose, as before, that $Z_{(1)}$ takes values $1, \ldots, J$ with probabilities $\theta_j^{(1)}$ where $\sum \theta_j^{(1)} = 1$. The model in Section 3 is generalized here to allow $Z_{(2)}$ to have an exponential family distribution of full rank. Thus, we suppose that the density of $Z_{(2)}$ given $Z_{(1)}$ is

$$f\big(Z_{(2)}|Z_{(1)} = j, \theta^{(2)}\big) = a\big(Z_{(2)}\big)\exp\big[c\big(\theta_j^{(2)}\big) + t\big(Z_{(2)}\big)^T\theta_j^{(2)}\big],$$

where $j = 1, \ldots, J$, $\theta_j^{(2)}$ and $t(Z_{(2)})$ are vectors of dimension V, and c is a real-valued function. This family includes the exponential and normal distribution (with variance known or unknown) as well as the multivariate normal, normal linear regression and generalized linear models with canonical links. The mean of $t(Z_{(2)})$ given $Z_{(1)} = j$ is given by the V-dimensional vector

$$\psi_j = \psi\big(\theta_j^{(2)}\big) = \frac{\partial c(\theta_j^{(2)})}{\partial \theta_j^{(2)}}.$$

In a random sample $(Z_{i(1)}, Z_{i(2)})$, $i = 1, \ldots, n$, the ML estimate of ψ_j is $\hat{\psi}_j = \sum t(Z_{i(2)})I(Z_{i(1)} = j)/n_{j+}$

TABLE 10
Estimates of girls' obesity rates

	Obesity rate				**Nuisance parameter**					**Observed data**
	θ_{11}	θ_{12}	θ_{21}	θ_{22}	ϕ	$\phi_1^{(0)}$	$\phi_2^{(0)}$	$\phi_1^{(1)}$	$\phi_2^{(1)}$	**loglikelihood**
Complete-case estimate	0.723	0.101	0.061	0.115	–	–	–	–	–	–
	(0.014)	(0.010)	(0.008)	(0.010)						
Restricted MAR								$\phi_1^{(1)} = \phi_2^{(1)}$		
					ϕ	$\phi_1^{(0)}$	$\phi_2^{(0)}$	$\phi^{(1)}$		
EM algorithm	0.685	0.099	0.073	0.143	0.304	0.383	0.518	0.241		
	(0.012)	(0.009)	(0.009)	(0.010)	(0.010)	(0.006)	(0.023)	(0.016)		−4535.605
Restricted BCMAR							$\phi_j^{(0)} = \phi_j^{(1)}$, $j = 1, 2$			
					ϕ	ϕ_1		ϕ_2		
EM algorithm	0.683	0.103	0.070	0.143	0.304	0.335		0.455		
	(0.011)	(0.009)	(0.008)	(0.010)	(0.010)	(0.006)		(0.022)		−4569.823
Unrestricted BCMAR					ϕ	$\phi_1^{(0)}$	$\phi_2^{(0)}$	$\phi_1^{(1)}$	$\phi_2^{(1)}$	
noniterative estimate	0.690	0.096	0.074	0.140	0.304	0.383	0.518	0.274	0.121	
	(0.012)	(0.010)	(0.009)	(0.010)	(0.010)	(0.006)	(0.023)	(0.034)	(0.122)	−4535.292

TABLE 11
Estimates of boys' obesity rates

	Obesity rate				Nuisance parameter					Observed data
	θ_{11}	θ_{12}	θ_{21}	θ_{22}	ϕ	$\phi_1^{(0)}$	$\phi_2^{(0)}$	$\phi_1^{(1)}$	$\phi_2^{(1)}$	loglikelihood
Complete-case estimate	0.710 (0.015)	0.099 (0.010)	0.073 (0.008)	0.118 (0.010)	–	–	–	–	–	–
Restricted MAR								$\phi_1^{(1)} = \phi_2^{(1)}$		
					ϕ	$\phi_1^{(0)}$	$\phi_2^{(0)}$	$\phi^{(1)}$		
EM algorithm	0.709 (0.011)	0.097 (0.009)	0.075 (0.008)	0.118 (0.008)	0.319 (0.009)	0.415 (0.006)	0.429 (0.025)	0.247 (0.015)		−4713.142
Restricted BCMAR							$\phi_j^{(0)} = \phi_j^{(1)}, j = 1, 2$			
					ϕ	ϕ_1		ϕ_2		
EM algorithm	0.709 (0.011)	0.098 (0.009)	0.075 (0.008)	0.118 (0.008)	0.319 (0.009)	0.360 (0.005)		0.375 (0.023)		−4748.480
Unrestricted BCMAR					ϕ	$\phi_1^{(0)}$	$\phi_2^{(0)}$	$\phi_1^{(1)}$	$\phi_2^{(1)}$	
noniterative estimate	0.707 (0.013)	0.099 (0.009)	0.074 (0.008)	0.120 (0.009)	0.319 (0.009)	0.415 (0.006)	0.429 (0.025)	0.228 (0.037)	0.325 (0.153)	−4713.027

where n_{j+} is the number of observations with $Z_{(1)} = j$; the ML estimate of $\theta_j^{(1)}$ is $\hat{\theta}_j^{(1)} = n_{j+}/\sum n_{l+}$. The ML estimates of $\theta_j^{(2)}$ can be obtained from those for ψ_j.

We consider as before the missing data structure illustrated in Table 1 with missingness patterns P_r with n_r observations, for $r = 0, \ldots, 3$. The missingness parameters $\Phi = (\phi, \phi_j^{(0)}, \phi_j^{(1)}, j = 1, \ldots, J)$ are defined as before in (3.1). The parameters in the model are denoted by the triple $(\theta^{(1)}, \theta^{(2)}, \Phi)$.

In this case, the likelihood contributions in each cell from the (incomplete) data are as follows:

- For $i \in P_0$, the observed data are $Z_{i(1)}, Z_{i(2)}$, $M_{i(1)} = M_{i(2)} = 0$ and the likelihood contribution is proportional to

$$A_0\big(Z_{i(1)} = j, Z_{i(2)}; \theta^{(1)}, \theta^{(2)}, \Phi\big) = \theta_j^{(1)} \exp\big[c\big(\theta_j^{(2)}\big) + t\big(Z_{i(2)}\big)^T \theta_j^{(2)}\big](1-\phi) \cdot \big(1 - \phi_j^{(0)}\big).$$

- For $i \in P_1$, the observed data are $Z_{i(1)}, M_{i(1)} = 0, M_{i(2)} = 1$, and the likelihood contribution is proportional to

$$A_1\big(Z_{i(1)} = j; \theta^{(1)}, \theta^{(2)}, \Phi\big) = \theta_j^{(1)}(1-\phi)\phi_j^{(0)}.$$

- For $i \in P_2$, the observed data are $Z_{i(2)}, M_{i(1)} = 1, M_{i(2)} = 0$ and the likelihood contribution is proportional to

$$A_2\big(Z_{(2)}; \theta^{(1)}, \theta^{(2)}, \Phi\big) = \phi \sum_{j=1}^{J} \theta_j^{(1)} \exp\big[c\big(\theta_j^{(2)}\big) + t\big(Z_{(2)}\big)^T \theta_j^{(2)}\big] \cdot \big(1 - \phi_j^{(1)}\big).$$

- For $i \in P_3$, no elements of $Z_{(1)}$ or $Z_{(2)}$ are observed and the data comprise $M_{i(1)} = 1, M_{i(2)} = 1$. The likelihood contribution is proportional to

$$A_3\big(\theta^{(1)}, \theta^{(2)}, \Phi\big) = \phi \sum_{j=1}^{J} \theta_j^{(1)} \phi_j^{(1)}.$$

The full observed-data likelihood is then the product of such terms and can be written as $L = L_0 L_1 L_2 L_3$, where

$$L_0 = (1-\phi)^{n_0} \prod_{j=1}^{J} \big\{\big(\theta_j^{(1)}\big)^{n_{(0),j+}} \big(1 - \phi_j^{(0)}\big)^{n_{(0),j+}} \cdot \exp\big[c\big(\theta_j^{(2)}\big) + T_{0j}^T \theta_j^{(2)}\big]\big\},$$

$$L_1 = (1-\phi)^{n_1} \prod_{j=1}^{J} \big\{\big(\theta_j^{(1)}\big)^{n_{(1),j+}} \big(\phi_j^{(0)}\big)^{n_{(1),j+}}\big\},$$

$$L_2 = \phi^{n_2} \prod_{i \in P_2} \Bigg\{ \sum_{j=1}^{J} \theta_j^{(1)} \big(1 - \phi_j^{(1)}\big) \cdot \exp\big[c\big(\theta_j^{(2)}\big) + t\big(Z_{i(2)}\big)^T \theta_j^{(2)}\big] \Bigg\},$$

$$L_3 = \phi^{n_3}\left\{\sum_{j=1}^{J}\theta_j^{(1)}\phi_j^{(1)}\right\}^{n_3},$$

and $T_{0j} = \sum_{i\in P_0} t(Z_{i(2)})I(Z_{i(1)} = j)$.

An EM algorithm can readily be applied to maximize the observed-data likelihood. At the E step, the underlying complete data in patterns P_2 and P_3 can be replaced with their conditional expectations, whereas blocks P_0 and P_1 can be treated as complete data. Alternatively, all four patterns can be incorporated into the EM approach, with the complete data viewed as all the observations $Z_{i(1)}, Z_{i(2)}, i = 1, \ldots, n$. For the data in block $i \in P_2$, for example, the expectation step involves calculating

$$E\big[I\big(Z_{i(1)} = j\big)|Z_{i(2)}, M_{i(1)} = 1, M_{i(2)} = 0\big] = \frac{\theta_j^{(1)}(1-\phi_j^{(1)})\exp[c(\theta_j^{(2)}) + t(Z_{i(2)})^T\theta_j^{(2)}]}{\sum_{l=1}^{J}\theta_l^{(1)}(1-\phi_l^{(1)})\exp[c(\theta_l^{(2)}) + t(Z_{i(2)})^T\theta_l^{(2)}]}.$$

After missing data in each pattern are filled in from the E step, the M step computes the simple estimates given above for complete data.

As in the multinomial case, the block-monotone reduced ML estimates of the parameters $\theta_j^{(1)}, \theta_j^{(2)}$, $j = 1, \ldots, J$, are computed from patterns P_0, P_1, dropping the data from the other patterns. The corresponding block-monotone reduced likelihood of $\theta^{(1)}, \theta^{(2)}$ is

$$L_{\mathrm{bm}}(P_0, P_1) \propto L_0 \times L_1,$$

where the factors in the parameters Φ can be ignored in L_0, L_1. Unlike the multinomial case, these block-monotone reduced ML estimates are typically not full ML estimates, since there is information about the parameters $\theta_j^{(2)}$ in the excluded patterns.

8. DISCUSSION

Most of the work on MNAR mechanisms concerns selection or pattern-mixture models, and extensions to include latent random effects that are applicable to repeated-measures data (Little, 1995). In this article we consider block-sequential missing data models, where the variables in the data set are divided into subsets, and the joint distribution of these variables and their missing data indicators are factored as a sequence. A characteristic of this class is that distributions of variables and their missing data indicators are interleaved, and combinations of selection and pattern-mixture models can be developed within each block. Except for the work of Robins (1997), there appears to be very little existing literature on missing data mechanisms of this type.

Here we consider a class of block-sequential missing data which we call block-conditional MAR models, in which missingness in successive blocks is allowed to depend on observed variables in the block and both observed and unobserved data in earlier blocks. The proposed class is related to the models with 2 blocks described in Little and Zhang (2011), in the context of regression with missing data. A block-monotone reduced likelihood approach to estimating these models is described that yields consistent asymptotically normal estimates without specifying the distribution of the missing-data mechanism. We examined here the BCMAR model in some detail for the case of bivariate categorical data, and showed that maximization of the block-monotone reduced likelihood can yield fully efficient ML estimates, when associated estimates of parameters of the missing-data mechanism lie inside the parameter space. We also discussed more briefly the case where the variable in the second block comes from an exponential family, and inference based on the block-monotone reduced likelihood approach is not in general fully efficient. In future work we plan to study other BCMAR models involving more than two blocks, continuous and categorical variables and missing data within each block, and fully observed covariate information.

The BCMAR model discussed here is related to the "latent ignorable" missing data mechanisms proposed to model missing data in the presence of noncompliance with a treatment (Frangakis and Rubin, 1999; Peng, Little and Raghunathan, 2004). In these cases, there is a binary compliance variable that indicates whether an individual would comply with a treatment if assigned to it. In a clinical trial, this indicator is fully observed for individuals in the active treatment group, but is completely missing for individuals in the control group, since they do not have access to the active treatment. The latent ignorable model assumes MAR within subpopulations defined by the compliance indicator. Our BCMAR model, applied to that setting, generalizes this structure by allowing missing data for the stratifying variable.

The BCMAR model (1.5) is just one of many possible block-sequential missing-data models, obtained by placing restrictions on the parameters of the distributions in each block. Future work might consider properties of models obtained by imposing other parameter restrictions, based on plausible assumptions about the nature of the missing data.

ACKNOWLEDGMENTS

We appreciate the constructive comments of two referees and an associate editor which greatly improved the paper.

REFERENCES

BAKER, S. G. (1995). Marginal regression for repeated binary data with outcome subject to non-ignorable nonresponse. *Biometrics* **51** 1042–1052.

BAKER, S. and LAIRD, N. (1985). Categorical response subject to nonresponse. Technical Report, Dept. Biostatistics, Harvard School of Public Health, Boston, MA.

BIRMINGHAM, J. and FITZMAURICE, G. M. (2002). A pattern-mixture model for longitudinal binary responses with nonignorable nonresponse. *Biometrics* **58** 989–996. MR1945028

BISHOP, Y. M. M., FIENBERG, S. E. and HOLLAND, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. MIT Press, Cambridge, MA. MR0381130

CHEN, T. and FIENBERG, S. E. (1974). Two-dimensional contingency tables with both completely and partially classified data. *Biometrics* **30** 629–642. MR0403086

COX, D. R. (1975). Partial likelihood. *Biometrika* **62** 269–276. MR0400509

DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc. Ser. B*. **39** 1–38. MR0501537

EKHOLM, A. and SKINNER, C. (1998). The muscatine children's obesity data reanalysed using pattern mixture models. *J. Roy. Statist. Soc. Ser. C*. **47** 251–263.

FAY, R. E. (1986). Causal models for patterns of nonresponse. *J. Amer. Statist. Assoc*. **81** 354–365.

FRANGAKIS, C. E. and RUBIN, D. B. (1999). Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika* **86** 365–379. MR1705410

FUCHS, C. (1982). Maximum likelihood estimation and model selection in contingency tables with missing data. *J. Amer. Statist. Assoc*. **77** 270–278.

HARTLEY, H. O. (1958). Maximum likelihood estimation from incomplete data. *Biometrics* **14** 174–194.

HECKMAN, J. I. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables, and a simple estimator for such models. *Ann. Econ. Soc. Meas.* **5** 475–492.

LIPSITZ, S. R., IBRAHIM, J. G. and ZHAO, L. P. (1999). A weighted estimating equation for missing covariate data with properites similar to maximum likelihood. *J. Amer. Statist. Assoc*. **94** 1147–1160. MR1731479

LIPSITZ, S. R., PARZEN, M. and MOLENBERGHS, G. (1998). Obtaining the maximum likelihood estimates in incomplete R×C contingency tables using a Poisson generalized linear model. *J. Comput. Graph. Statist.* **7** 356–376.

LITTLE, R. J. A. (1993). Pattern-mixture models for multivariate incomplete data. *J. Amer. Statist. Assoc*. **88** 125–134.

LITTLE, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *J. Amer. Statist. Assoc*. **90** 1112–1121. MR1354029

LITTLE, R. J. A. and RUBIN, D. B. (2002). *Statistical Analysis with Missing Data.* Wiley, Hoboken, NJ. MR1925014

LITTLE, R. J. A. and ZHANG, N. (2011). Subsample ignorable likelihood for regression analysis with missing data. *J. Roy. Statist. Soc. Ser. C* **60**. To appear.

MENG, X.-L. and RUBIN, D. B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika* **80** 267–278. MR1243503

PENG, Y. H., LITTLE, R. J. A. and RAGHUNATHAN, T. E. (2004). An extended general location model for causal inferences from data subject to noncompliance and missing values. *Biometrics* **60** 598–607. MR2089434

ROBINS, J. M. (1997). Non-response models for the analysis of non-monotone non-ignorable missing data. *Stat. Med.* **16** 21–37.

ROBINS, J. M. and GILL, R. (1997). Non-response models for the analysis of non-monotone ignorable missing data. *Stat. Med.* **16** 39–56.

RUBIN, D. B. (1976). Inference and missing data (with discussion). *Biometrika* **63** 581–592. MR0455196

RUBIN, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys.* Wiley, New York. MR0899519

RUBIN, D. B., STERN, H. and VEHOVAR, V. (1995). Handling "don't know" survey responses: The case of the Slovenian plebiscite. *J. Amer. Statist. Assoc*. **90** 822–828.

WOOLSON, R. F. and CLARKE, W. R. (1984). Analysis of categorical incomplete longitudinal data. *J. Roy. Statist. Soc. Ser. A* **147** 87–99.