Computational Statistics and Data Analysis 75 (2014) 239–247

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

A hybrid approach for regression analysis with block missing data

Zhengbang Li [a,*], Qizhai Li [b], Chien-Pai Han [c], Bo Li [a]

[a] School of Mathematics and Statistics, Central China Normal University, Wuhan, China

[b] Academy of Mathematics and Systems Science, Chinese Academy of Sciences, China

[c] Department of Mathematics, University of Texas at Arlington, TX, USA

ARTICLE INFO

Article history:
Received 7 October 2012
Received in revised form 12 February 2014
Accepted 14 February 2014
Available online 22 February 2014

Keywords:
Block missing
Hybrid
Likelihood
Bayesian
Frequentist

ABSTRACT

Missing data often arise in practice. The commonly employed approach to handle the missing data is imputation, which is effective when the missing mechanism is known and each subject in the data set misses at random. However, the situation where the imputation is not appropriate often emerged. Because in that situation, some data are not missing at random, so a hybrid estimate, where the Bayesian and frequentist approaches are used for inferring the parameters with and without prior information respectively, is proposed. The asymptotic properties of the hybrid estimator are also provided. Numerical results including simulation studies and data analysis about grade point average (GPA) are conducted to show the performances of the proposed method.

1. Introduction

Missing data is encountered frequently in practice. For example, in a study to examine whether the undergraduate grade point average (GPA), graduate record examination (GRE) verbal, GRE quantitative, GRE analytic, and the test of English as a foreign language (TOEFL) are related to the master students's GPA in the United States (US) universities. Admission office at the University of Texas at Arlington collected the data in the past years. The data contains two groups where one group consists of the students from the outside-US countries, another is from US. As well known that the students from other countries, especially from the non-English spoken countries, have to take TOEFL, and thus they have the TOEFL records. The US students do not need to take the TOEFL, and they have no TOEFL scores. In this case the TOEFL scores of all US students are missing. The question is how to do reasonable inference on the relationship between the master students's GPA and the independent variables of interest, including TOFEL scores from students of non-English spoken countries.

A lot of work has been done in the area of missing by now. For example, Anderson (1957), Nicholson (1957), Hocking and Smith (1968), and Ratkowsky (1974) discussed the incomplete samples for the multivariate normal cases; Sundberg (1974), Dempster et al. (1977), Dahiya and Korwar (1980), Rochon and Helms (1984), and Lipsitz et al. (1999) considered using the likelihood method to handle the missing data; Titterington (1984) and Sexton and Laake (2009) proposed to use the stochastic approximation procedures to estimate the parameter of interest for incomplete data; Woolson and Clarke (1984) considered the analysis for incomplete longitudinal data. Chen and Little (1999) considered the testing problem for generalized estimating equations with missing data, etc. There are also some review work; see Afifi and Elashoff (1966), Rubin (1976, 1987), and Little and Rubin (2000) etc.

* Correspondence to: School of Mathematics and Statistics, CCNU, No. 152, Luoyu Road, Wuhan, China. Tel.: +86 027 67867450; fax: +86 027 67867450.
E-mail address: lizhengbang@mail.ccnu.edu.cn (Z. Li).

http://dx.doi.org/10.1016/j.csda.2014.02.014
0167-9473/

Although many procedures are available for handling the missing data, the most popular technique is imputation, which is based on the assumptions that the distribution of missing data is the same as the observed ones, and each subject in the data has the equal missing probability. In GPA data, the US students do not need to take TOEFL, and thus there is "no" distribution for the TOEFL scores of US students. So the imputation for missing TOEFL scores of US students is not appropriate. Recently, Han and Li (2011) proposed a procedure that combined two estimators, one is obtained based on the observations with missing variables, and the other is gotten without considering missing variables, where the missing variable is defined to be the variable with its data missing. Their procedure is different from imputation. In this work, we propose an alternative estimator for estimating the regression coefficients. When the observations for some covariates are complete, we use a Bayesian inference, and for the parameter with missing data, we employ a frequentist method. According to Yuan (2009), we call it a hybrid inference. The asymptotic properties of the proposed estimator are derived under some regular conditions for the 0–1 loss function. Prior distributions of parameters are estimated using the existing data. Yuan and De Gooijer (in press) introduced the asymptotically informative prior for the Bayesian analysis, which supplies us insightful ideas for the hybrid inference. Our method can be easily adapted in the case of both known and asymptotically informative prior distributions of parameters.

This paper is organized as follows. In Section 2, we give the notations and the proposed method with its asymptotic statistical properties. In Section 3, we compare several estimators using simulation studies and the GPA data analysis. Some discussions are given at last.

2. Methods

2.1. Notations

Consider the linear regression model

$$Y = X^\tau \beta + Z^\tau \gamma + \varepsilon, \tag{1}$$

where $X = (X_1, \dots, X_k)^\tau$, $Z = (Z_1, \dots, Z_r)^\tau$ are independent variables with $X_1 = 1$ being for the intercept, $\beta = (\beta_1, \dots, \beta_k)^\tau$ and $\gamma = (\gamma_1, \dots, \gamma_r)^\tau$ are parameters, and $\varepsilon \sim N(0, \sigma^2)$, σ^2 is unknown.

Suppose that $m + n$ subjects are sampled in a general population, where n subjects have no observations on Z. Denote the ith observation by $(y_i, x_{i1}, \dots, x_{ik}, z_{i1}, \dots, z_{ir})$, $i = 1, \dots, m$, and $(y_i, x_{i1}, \dots, x_{ik})$, $i = m+1, \dots, m+n$. The data are, in the matrix notation, $\mathbf{Y}$, $\mathbf{X}$ and $\mathbf{Z}_+$, where $\mathbf{Y} = \left(\mathbf{Y}_+^\tau, \mathbf{Y}_-^\tau\right)^\tau$, $\mathbf{Y}_+ = (y_1, \dots, y_m)^\tau$, $\mathbf{Y}_- = (y_{m+1}, \dots, y_{m+n})^\tau$, $\mathbf{X} = (x_{ij})_{(m+n)\times k} = \left(\mathbf{X}_+^\tau, \mathbf{X}_-^\tau\right)^\tau$, $\mathbf{X}_+ = (x_{ij})_{m\times k}$, and $\mathbf{Z}_+ = (z_{ij})_{m\times r}$.

2.2. Inference of the parameters

Without considering the missing data, an ad hoc approach to estimate β and γ is the ordinary least square approach. When the random error follows the normal distribution, the estimates from ordinary least square are equivalent to the maximum likelihood estimates. The likelihood function is

$$L(\beta, \gamma, \sigma^2) = (2\pi\sigma^2)^{-\frac{m}{2}} \exp\left\{-\frac{(\mathbf{Y}_+ - \mathbf{X}_+\beta - \mathbf{Z}_+\gamma)^\tau(\mathbf{Y}_+ - \mathbf{X}_+\beta - \mathbf{Z}_+\gamma)}{2\sigma^2}\right\}, \tag{2}$$

and the log-likelihood function is

$$l(\beta, \gamma, \sigma^2) = -\frac{m}{2}(\log 2\pi + \log\sigma^2) - \frac{(\mathbf{Y}_+ - \mathbf{X}_+\beta - \mathbf{Z}_+\gamma)^\tau(\mathbf{Y}_+ - \mathbf{X}_+\beta - \mathbf{Z}_+\gamma)}{2\sigma^2}. \tag{3}$$

If we have some prior information on β, not on γ, the classical Bayesian analysis might not be applied directly. One may put some non-informative prior on γ, to formulate a full Bayesian model. However, such prior may mislead the results more or less, and it is known that a full Bayesian method is computationally much more demanding than the corresponding frequentist method. On the other hand, if a full frequentist method is used, the computation is simple, but the prior information on β will be wasted. In the following, we will show that the prior information on β can be implemented, and the computation is parallel to that of the MLE. Here we adopt the idea of Yuan (2009) and propose to use a hybrid estimate to infer θ where $\theta = (\beta^\tau, \gamma^\tau, \sigma^2)^\tau$. Denote the prior density of β by $\pi(\beta)$. Let $w(d, \beta)$ be the loss function, and $d = d(\mathbf{Y}_+|\mathbf{X}_+; \mathbf{Z}_+)$ be the decision for β. Denote the parameter spaces for β, γ and σ^2 by B, Γ and R^+ respectively. Then, the hybrid estimate (θ^*) of θ, $\theta^* = \left((\beta^*)^\tau, (\gamma^*)^\tau, (\sigma^2)^*\right)^\tau$ is calculated as follows: for the fixed γ and σ^2, β^* is obtained using the Bayesian rule, which is

$$\beta^* = \operatorname*{arginf}_{d\in B} \int_B w(d, \beta) f(\mathbf{Y}_+|\mathbf{X}_+, \mathbf{Z}_+, \theta)\pi(\beta)d\beta, \tag{4}$$

where $f(\cdot)$ denotes the density function; for a fixed β, γ^* and $(\sigma^2)^*$ are the MLEs of γ and σ^2, respectively.

There are three often used loss functions including the quadratic loss, absolute loss and the 0–1 loss functions. Yuan (2009) discussed the 3 loss functions. Comparing between these three loss functions, the 0–1 loss function could simplify the computation since the hybrid estimate is similar to the usual MLE. When the 0–1 loss is employed,

$$\theta^* = \arg \sup_{\gamma\in\Gamma,\sigma^2\in R^+} \sup_{\beta\in B} \left(l(\beta,\gamma,\sigma^2) + \ln\pi(\beta)\right). \tag{5}$$

How to chose $\pi(\beta)$? A good alternative is to use $\mathbf{X}_-$ to determine $\pi(\beta)$. We regress $\mathbf{Y}_-$ on $\mathbf{X}_-$, and thus an estimator of β is given by $\tilde{\beta} = (\mathbf{X}_-^\tau\mathbf{X}_-)^{-1}\mathbf{X}_-^\tau\mathbf{Y}_-$, with the variance being $\mathrm{var}(\tilde{\beta}) = \sigma^2(\mathbf{X}_-^\tau\mathbf{X}_-)^{-1}$. We take the density of $N((\mathbf{X}_-^\tau\mathbf{X}_-)^{-1}\mathbf{X}_-^\tau\mathbf{Y}_-,\ \sigma^2(\mathbf{X}_-^\tau\mathbf{X}_-)^{-1})$ as $\pi(\beta)$, where we keep σ^2 in $\mathrm{var}(\tilde{\beta})$ although it is unknown. In practice, σ^2 will vanish in the derivation of β^* and γ^*. Denote

$$\begin{aligned} l^*(\beta,\gamma,\sigma^2) &= l(\beta,\gamma,\sigma^2) + \ln[\pi(\beta)] \\ &= -\frac{m}{2}(\ln 2\pi + \ln\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{m}(Y_i - X_i{}^\tau\beta - Z_i{}^\tau\gamma)^2 \\ &\quad - \frac{k}{2}(\ln 2\pi + \log\sigma^2) - \frac{\ln|\mathbf{X}_-^\tau\mathbf{X}_-|}{2} - \frac{(\beta-\tilde{\beta})^\tau\mathbf{X}_-^\tau\mathbf{X}_-(\beta-\tilde{\beta})}{2\sigma^2}, \end{aligned} \tag{6}$$

which is called the hybrid likelihood function. Then, the normal equations are

$$\begin{cases} \dfrac{\partial}{\partial\beta}l^*(\beta,\gamma,\sigma^2) = \dfrac{\sum_{i=1}^{m}X_i(Y_i - X_i{}^\tau\beta - Z_i{}^\tau\gamma)}{\sigma^2} - \dfrac{\mathbf{X}_-^\tau\mathbf{X}_-(\beta-\tilde{\beta})}{\sigma^2} = 0 \\ \dfrac{\partial}{\partial\gamma}l^*(\beta,\gamma,\sigma^2) = \dfrac{\partial}{\partial\gamma}l(\beta,\gamma,\sigma^2) = \dfrac{\sum_{i=1}^{m}Z_i(Y_i - X_i{}^\tau\beta - Z_i{}^\tau\gamma)}{\sigma^2} = 0 \\ \dfrac{\partial}{\partial(\sigma^2)}l^*(\beta,\gamma,\sigma^2) = \dfrac{\sum_{i=1}^{m}(Y_i - X_i\beta - Z_i\gamma)^2}{2\left(\sigma^2\right)^2} - \dfrac{m}{2\sigma^2} + \dfrac{(\beta-\tilde{\beta})^\tau\mathbf{X}_-^\tau\mathbf{X}_-(\beta-\tilde{\beta})}{2\left(\sigma^2\right)^2} - \dfrac{k}{2\sigma^2} = 0. \end{cases} \tag{7}$$

In the matrix notation, we have

$$\begin{cases} \dfrac{\mathbf{X}_+{}^\tau\mathbf{Y}_+ - \mathbf{X}_+{}^\tau\mathbf{X}_+\beta - \mathbf{X}_+{}^\tau\mathbf{Z}_+\gamma}{\sigma^2} - \dfrac{\mathbf{X}_-^\tau\mathbf{X}_-(\beta-\tilde{\beta})}{\sigma^2} = 0 \\ \dfrac{\mathbf{Z}_+{}^\tau\mathbf{Y}_+ - \mathbf{Z}_+{}^\tau\mathbf{X}_+\beta - \mathbf{Z}_+{}^\tau\mathbf{Z}_+\gamma}{\sigma^2} = 0 \\ (\mathbf{Y}_+ - \mathbf{X}_+\beta - \mathbf{Z}_+\gamma)^\tau(\mathbf{Y}_+ - \mathbf{X}_+\beta - \mathbf{Z}_+\gamma) + (\beta-\tilde{\beta})^\tau\mathbf{X}_-^\tau\mathbf{X}_-(\beta-\tilde{\beta}) = (m+k)\sigma^2. \end{cases} \tag{8}$$

After some algebra, we have

$$\begin{cases} \beta^* = V^{-1}\left[\mathbf{X}_+{}^\tau\mathbf{Z}_+(\mathbf{Z}_+{}^\tau\mathbf{Z}_+)^{-1}\mathbf{Z}_+{}^\tau\mathbf{Y}_+ - (\mathbf{X}_+{}^\tau\mathbf{Y}_+ + \mathbf{X}_-{}^\tau\mathbf{X}_-\tilde{\beta})\right] \\ \gamma^* = \left(\mathbf{Z}_+{}^\tau\mathbf{Z}_+\right)^{-1}\left(\mathbf{Z}_+{}^\tau\mathbf{Y}_+ - \mathbf{Z}_+{}^\tau\mathbf{X}_+\beta^*\right) \\ (\sigma^2)^* = \dfrac{(\mathbf{Y}_+ - \mathbf{X}_+\beta^* - \mathbf{Z}_+\gamma^*)^\tau(\mathbf{Y}_+ - \mathbf{X}_+\beta^* - \mathbf{Z}_+\gamma^*) + (\beta^*-\tilde{\beta})^\tau\mathbf{X}_-^\tau\mathbf{X}_-(\beta^*-\tilde{\beta})}{m+k}, \end{cases} \tag{9}$$

where $V = \mathbf{X}_+{}^\tau\mathbf{Z}_+(\mathbf{Z}_+{}^\tau\mathbf{Z}_+)^{-1}\mathbf{Z}_+{}^\tau\mathbf{X}_+ - (\mathbf{X}_+{}^\tau\mathbf{X}_+ + \mathbf{X}_-{}^\tau\mathbf{X}_-)$, $\tilde{\beta} = (\mathbf{X}_-^\tau\mathbf{X}_-)^{-1}\mathbf{X}_-^\tau\mathbf{Y}_-$.

The asymptotic properties of θ^* are given in Theorem 1.

Theorem 1. *Let* $\mathring{\theta} = \left((\mathring{\beta})^\tau, (\mathring{\gamma})^\tau, (\mathring{\sigma})^2\right)^\tau$ *be the true value of* θ. *Suppose that* $\max\{k, r\} < \infty$, *and* $\lim_{m\to\infty}\frac{1}{m}\begin{pmatrix}\mathbf{X}_+^\tau\mathbf{X}_+ & \mathbf{X}_+^\tau\mathbf{Z}_+ \\ \mathbf{Z}_+^\tau\mathbf{X}_+ & \mathbf{Z}_+^\tau\mathbf{Z}_+\end{pmatrix} = \mathbf{W}_1$.

(1) *Let n be fixed. Under the* 0–1 *loss, as* $m\to\infty$, *we have*

$$\text{(i) } \theta^* \xrightarrow{\text{a.s.}} \mathring{\theta}. \qquad \text{(ii) } \sqrt{m}\left((\beta^*)^\tau - (\mathring{\beta})^\tau, (\gamma^*)^\tau - (\mathring{\gamma})^\tau\right)^\tau \xrightarrow{D} N\left(0, (\mathring{\sigma})^2\mathbf{W}_1^{-1}\right).$$

(2) *Suppose that* $\lim_{m,n\to\infty}\frac{\mathbf{X}_-^\tau\mathbf{X}_-}{m} = \mathbf{W}_2$, *and m grows faster than n so that* $\lim_{m,n\to\infty}\frac{\mathbf{X}_-^\tau}{m} = \mathbf{0}_{k\times n}$ *(matrix with all elements being 0). Denote* $\mathbf{W} = \mathbf{W}_1 + \begin{pmatrix}\mathbf{W}_2 & \mathbf{0}_{k\times r} \\ \mathbf{0}_{r\times k} & \mathbf{0}_{r\times r}\end{pmatrix}$, *where* $\mathbf{0}_{k\times r}, \mathbf{0}_{r\times k}$ *and* $\mathbf{0}_{r\times r}$ *are matrices with all elements being* 0. *Then, under* 0–1 *loss, as* $m, n\to\infty$, *we can get:*

$$\text{(iii) } \theta^* \xrightarrow{\text{a.s.}} \mathring{\theta}. \qquad \text{(iv) } \sqrt{m}\left((\beta^*)^\tau - (\mathring{\beta})^\tau, (\gamma^*)^\tau - (\mathring{\gamma})^\tau\right)^\tau \xrightarrow{D} N\left(0, (\mathring{\sigma})^2\mathbf{W}^{-1}\right).$$

Table 1
Empirical biases of HE, OLS and CE.

	True	HE $m = 4000$ $n = 1000$	HE $m = 2500$ $n = 2500$	HE $m = 1000$ $n = 4000$	True	HE $m = 4000$ $n = 1000$	HE $m = 2500$ $n = 2500$	HE $m = 1000$ $n = 4000$
β_1	0.5	0.0005	0.0002	−0.0009	0.5	−0.0021	0.0008	0.0018
β_2	1	−0.0005	−0.0003	−0.0001	1	0.0001	0.0001	0.0031
β_3	1	0.0003	0.0008	−0.0001	1	−0.0011	0.0014	0.0014
β_4	1	−0.0004	0.0009	−0.0007	1	−0.0002	0.0008	−0.0018
β_5	1	−0.0004	−0.0004	−0.0002	1	−0.0006	−0.0009	−0.0005
γ_1	0	−0.0001	0.005	−0.0006	0.2	−0.0002	−0.0005	−0.0006
β_1	0.5	−0.0015	0.0030	0.0028	0.5	−0.0091	0.0042	−0.0042
β_2	1	0.0005	−0.0004	0.0098	1	−0.0002	−0.0000	0.0092
β_3	1	−0.0017	0.0018	0.0029	1	−0.0041	0.0029	0.0047
β_4	1	−0.0013	0.0023	−0.0036	1	−0.0016	0.0037	−0.0066
β_5	1	−0.0018	−0.0031	−0.0003	1	−0.0026	−0.0054	−0.0011
γ_1	0.5	−0.0001	−0.0007	0.0002	0.8	0.0006	−0.0001	−0.0004
β_1	0.5	−0.0013	0.0055	0.0051	0.5	0.0016	0.0065	0.0063
β_2	1	0.0001	−0.0006	0.0102	1	−0.0008	0.0001	0.0142
β_3	1	−0.0053	0.0042	0.0069	1	−0.0059	0.0045	0.0072
β_4	1	−0.0021	0.0042	−0.0076	1	−0.0031	0.0045	−0.0094
β_5	1	−0.0036	−0.0071	−0.0017	1	−0.0043	−0.0082	−0.0012
γ_1	1	0.0009	−0.0003	0.0004	1.2	−0.0006	−0.0003	−0.0037

The proof of Theorem 1 is given in the Appendix. As well known, in the classical Bayesian framework, the prior is fixed, under broad conditions, both the Bayesian estimator and the MLE are efficient. From the proof of Theorem 1, we can see that the asymptotic variance–covariance matrix W_1 is exact the inverse of Fisher information lower bound. When n is fixed as m goes to infinity, our hybrid estimate is asymptotically efficient.

We can also see from Theorem 1, when the covariates **X** and **Z** obey some reasonable conditions, although both n and m go to infinity, we still have the unbiased parameter estimates. This is verified by simulation studies in the next section.

3. Numerical results

3.1. Simulation studies

In this section, we conduct simulation studies to evaluate the performances of the proposed hybrid estimator (HE) compared with the ordinary least square estimator (OLS) and the combined estimators (CE) proposed by Han and Li (2011).

In order to verify the asymptotic unbias property of our estimate, we first consider 3 scenarios: $m = 4000$ and $n = 1000$, $m = 2500$ and $n = 2500$, and $m = 1000$ and $n = 4000$. We assume that the random errors are independently drawn from the standard normal distribution. Because X_2, X_3, X_4, X_5 and Z_1 are fixed effects, we generate them from the 5-dimensional normal distribution with zero mean and variance–covariance matrix $\Sigma = I_5$, the 5×5 identity matrix, and fix them during the replication process. 1000 replicates are conducted to calculate the empirical biases. Table 1 shows the empirical results of biases, which indicates that the proposed HE is asymptotically unbiased where the true values are: $\beta_1 = 0.5$, $\beta_2 = 1$, $\beta_3 = 1$, $\beta_4 = 1$, $\beta_5 = 1$, and $\gamma_1 \in \{0, 0.2, 0.5, 0.8, 1, 1.2\}$.

Since in GPA data, $k = 5$ and $r = 1$, we consider $k = 5$ and $r = 1$ in the simulations. We set $\beta_1 = 0.5$, $\beta_2 = 1$, $\beta_3 = 1$, $\beta_4 = 1$, $\beta_5 = 1$, and $\gamma_1 \in \{0, 0.2, 0.5, 0.8, 1, 1.2\}$. Three scenarios are considered: $m = 40$ and $n = 10$, $m = 25$ and $n = 25$, and $m = 10$ and $n = 40$. We assume that the random errors are independently drawn from the standard normal distribution. Because X_2, X_3, X_4, X_5 and Z_1 are fixed effects, we generate them from the 5-dimensional normal distribution with the zero mean vector and variance–covariance matrix $\Sigma = I_5$, the 5×5 identity matrix, and also fix them during the replication process. 10,000 replicates are conducted to calculate the empirical bias and mean squared error (MSE). Table 2 shows the empirical biases for $m = 40$ and $n = 10$; Table 3 shows the empirical MSEs for $m = 40$ and $n = 10$; Table 4 shows the empirical biases for $m = 25$ and $n = 25$; Table 5 shows the empirical MSEs for $m = 25$ and $n = 25$; Table 6 shows the empirical biases for $m = 10$ and $n = 40$; Table 7 shows the empirical MSEs for $m = 10$ and $n = 40$.

From the tables, we find that the prior information for β depends on n, the number of observations for subjects with missing covariates. When both m and n are not very large, the empirical biases are not ignorable. Obviously, if n is larger than m (the number of subjects with complete observations), there appears to be large biases. We also find that the biases of the proposed HE are larger than those of OLS, which is reasonable since OLS estimate is unbiased. The biases of HE and CE are comparable, while the MSEs of HE are always smaller than those of OLS and CE over all. When $\gamma_1 = 0$, the empirical bias and MSE of HE are smaller than those of OLS and CE. This is reasonable since when $\gamma_1 = 0$, HE is equivalent to OLS and more samples are utilized by HE. When γ_1 becomes large such as 0.8, 1.0, and 1.2, although the empirical bias of HE is larger than that of OLS, the empirical MSE is much smaller that of OLS. For example, when $\gamma_1 = 0.8$ the empirical biases of HE, OLS and CE for γ_1 are −0.1301, 0.0002, and 0.0002, respectively, while the MSE of HE is 0.1445, which is almost one-third of those for OLS and CE (both MSEs are equal to 0.3832).

Table 2
Empirical biases of HE, OLS and CE ($m = 40$, $n = 10$).

	True	HE	OLS	CE	True	HE	OLS	CE
β_1	0.5	−0.0003	−0.0003	−0.0003	0.5	−0.0041	0.0025	−0.0004
β_2	1	−0.0003	0.0004	−0.0002	1	0.0078	−0.0009	0.0135
β_3	1	0.0002	0.0008	0.0004	1	0.0222	0.0014	0.0151
β_4	1	0.0008	−0.0002	−0.0006	1	−0.0065	0.0001	−0.0109
β_5	1	−0.0001	−0.0005	−0.0003	1	0.0186	−0.0010	0.0201
γ_1	0	−0.002	−0.002	−0.002	0.2	0.0001	0.0016	0.0016
β_1	0.5	−0.0127	0.0015	−0.0040	0.5	−0.0222	−0.0008	−0.0086
β_2	1	0.0175	−0.0053	0.0313	1	0.0344	−0.0013	0.0566
β_3	1	0.0486	−0.0003	0.0316	1	0.0814	0.0002	0.0533
β_4	1	−0.0166	−0.0011	−0.0278	1	−0.0248	−0.0015	−0.0431
β_5	1	0.0530	0.0021	0.0561	1	0.0806	−0.0005	−0.0856
γ_1	0.5	−0.0015	0.0024	0.0024	0.8	−0.0035	0.0026	0.0026
β_1	0.5	−0.0265	0.0026	−0.0090	0.5	−0.0332	−0.0001	−0.0126
β_2	1	0.0429	−0.0007	0.0709	1	0.0523	0.0003	0.0858
β_3	1	0.1018	−0.0001	0.0666	1	0.1200	−0.0010	0.0781
β_4	1	−0.0314	−0.0016	−0.0541	1	−0.0392	−0.0031	−0.0663
β_5	1	0.1008	0.0013	0.1075	1	0.1217	0.0018	0.1295
γ_1	1	−0.0065	0.0011	0.0011	1.2	−0.0098	−0.0007	−0.0007

Table 3
Empirical MSEs of HE, OLS and CE ($m = 40$, $n = 10$).

	True	HE	OLS	CE	True	HE	OLS	CE
β_1	0.5	0.0262	0.0334	0.0268	0.5	0.0262	0.0336	0.0268
β_2	1	0.0367	0.0403	0.0369	1	0.0387	0.0422	0.0390
β_3	1	0.0222	0.0301	0.0228	1	0.0229	0.0304	0.0231
β_4	1	0.0201	0.0248	0.0204	1	0.0201	0.0249	0.0205
β_5	1	0.0242	0.0307	0.0246	1	0.0251	0.0315	0.0256
γ_1	0	0.0302	0.0303	0.0303	0.2	0.0299	0.0300	0.0300
β_1	0.5	0.0266	0.0338	0.0270	0.5	0.0274	0.0338	0.0275
β_2	1	0.0377	0.0404	0.0384	1	0.0399	0.0422	0.0421
β_3	1	0.0243	0.0298	0.0234	1	0.0291	0.0302	0.0257
β_4	1	0.0200	0.0248	0.0208	1	0.0207	0.0251	0.0222
β_5	1	0.0272	0.0311	0.0280	1	0.0309	0.0306	0.0321
γ_1	0.5	0.0311	0.0312	0.0312	0.8	0.0304	0.0305	0.0305
β_1	0.5	0.0275	0.0344	0.0275	0.5	0.0272	0.0330	0.0267
β_2	1	0.0404	0.0418	0.0437	1	0.0410	0.0416	0.0457
β_3	1	0.0323	0.0300	0.0268	1	0.0366	0.0304	0.0289
β_4	1	0.0205	0.0244	0.0228	1	0.0211	0.0241	0.0241
β_5	1	0.0346	0.0309	0.0364	1	0.0391	0.0306	0.0414
γ_1	1	0.0311	0.0311	0.0311	1.2	0.0308	0.0308	0.0308

Table 4
Empirical biases of HE, OLS and CE ($m = 25$, $n = 25$).

	True	HE	OLS	CE	True	HE	OLS	CE
β_1	0.5	0.0001	−0.0024	0.0002	0.5	−0.0311	−0.0025	−0.0434
β_2	1	0.0016	−0.0035	0.0015	1	−0.0403	−0.0042	−0.0328
β_3	1	−0.0000	−0.0080	0.0001	1	−0.0367	−0.0006	−0.0384
β_4	1	−0.0003	−0.0021	−0.0005	1	−0.0059	−0.0010	0.0069
β_5	1	0.0001	−0.0030	0.0001	1	−0.0146	−0.0008	−0.0305
γ_1	0	−0.0016	−0.0027	−0.0027	0.2	−0.0147	−0.0046	−0.0046
β_1	0.5	−0.0746	−0.0031	−0.1062	0.5	−0.1214	0.0001	−0.1719
β_2	1	−0.0942	0.0014	−0.0750	1	−0.1502	0.0025	−0.1193
β_3	1	−0.0991	−0.0004	−0.1033	1	−0.1603	−0.0026	−0.1670
β_4	1	−0.0194	−0.0007	0.0135	1	−0.0310	0.0016	0.0217
β_5	1	−0.0409	0.0010	−0.0815	1	−0.0638	−0.0001	−0.1289
γ_1	0.5	−0.0267	−0.0011	−0.0011	0.8	−0.0418	−0.0003	−0.0003
β_1	0.5	−0.1505	0.0040	−0.2136	0.5	−0.1861	−0.0022	−0.2613
β_2	1	−0.1934	−0.0038	−0.1548	1	0.2315	−0.0055	−0.1854
β_3	1	−0.1988	−0.0025	−0.2071	1	−0.2400	−0.0035	−0.2500
β_4	1	−0.0369	0.0025	0.0291	1	−0.0454	−0.0005	0.0333
β_5	1	−0.0805	−0.0020	−0.1619	1	−0.0961	−0.0042	−0.1932
γ_1	1	−0.0512	0.0016	0.0016	1.2	−0.0683	−0.0052	−0.005

Table 5
Empirical MSEs of HE, OLS and CE ($m = 25$, $n = 25$).

	True	HE	OLS	CE	True	HE	OLS	CE
β_1	0.5	0.0272	0.0481	0.0326	0.5	0.0226	0.0492	0.0231
β_2	1	0.0415	0.0860	0.0448	1	0.0407	0.0852	0.0401
β_3	1	0.0488	0.1159	0.0497	1	0.0413	0.1160	0.0414
β_4	1	0.0230	0.0525	0.0225	1	0.0224	0.0525	0.0221
β_5	1	0.0294	0.0482	0.0338	1	0.0278	0.0474	0.0279
γ_1	0	0.0766	0.0813	0.0813	0.2	0.0781	0.0836	0.0836
β_1	0.5	0.0247	0.2849	0.0270	0.5	0.0367	0.0480	0.0512
β_2	1	0.0208	0.3144	0.0219	1	0.0618	0.0849	0.0533
β_3	1	0.0359	0.2381	0.0344	1	0.0653	0.1167	0.0675
β_4	1	0.0427	0.2166	0.0510	1	0.0236	0.0540	0.0228
β_5	1	0.0275	0.1652	0.0374	1	0.0318	0.0481	0.0438
γ_1	0.5	0.1341	0.3900	0.3900	0.8	0.0770	0.0809	0.0809
β_1	0.5	0.0448	0.0497	0.0674	0.5	0.0559	0.0494	0.0893
β_2	1	0.0770	0.0860	0.0635	1	0.0935	0.0879	0.0743
β_3	1	0.0794	0.1165	0.0829	1	0.0987	0.1196	0.1036
β_4	1	0.0235	0.0523	0.0227	1	0.0249	0.0532	0.0236
β_5	1	0.0341	0.0483	0.0533	1	0.0373	0.0485	0.0648
γ_1	1	0.0787	0.0816	0.0816	1.2	0.0804	0.0811	0.0811

Table 6
Empirical biases of HE, OLS and CE ($m = 10$, $n = 40$).

	True	HE	OLS	CE	True	HE	OLS	CE
β_1	0.5	−0.0010	−0.0028	−0.0002	0.5	−0.0055	−0.0030	0.0314
β_2	1	−0.0004	−0.0028	0.0010	1	0.0059	−0.0063	0.0197
β_3	1	0.0012	−0.0025	0.0012	1	−0.0420	−0.0057	−0.0389
β_4	1	0.0009	0.0055	0.0004	1	0.0453	0.0084	0.0593
β_5	1	−0.0001	0.0022	0.0006	1	−0.0304	−0.0021	−0.0531
γ_1	0	−0.0058	−0.0051	−0.0051	0.2	−0.0353	−0.0054	−0.0054
β_1	0.5	−0.0103	0.0016	−0.0754	0.5	−0.0159	0.0059	−0.1210
β_2	1	0.0113	0.0012	0.0461	1	0.0217	−0.0022	0.0778
β_3	1	−0.1047	0.0112	−0.0969	1	−0.1678	−0.0001	−0.1553
β_4	1	0.1128	−0.0071	0.1480	1	0.1824	0.0033	0.2392
β_5	1	−0.0809	0.0035	−0.1381	1	−0.1280	0.0015	−0.2202
γ_1	0.5	−0.0851	−0.0036	−0.0036	0.8	−0.1301	0.0002	0.0002
β_1	0.5	−0.0190	0.0065	−0.1514	0.5	−0.0251	−0.0059	−0.1837
β_2	1	0.0255	0.0019	0.0961	1	0.0320	0.0056	0.1166
β_3	1	−0.2108	−0.0110	−0.1950	1	−0.2518	−0.0002	−0.2328
β_4	1	0.2226	−0.0011	0.2942	1	0.2751	−0.0004	0.3607
β_5	1	−0.1598	−0.0014	0.2760	1	−0.1919	0.0021	−0.3311
γ_1	1	−0.1567	−0.0052	−0.0052	1.2	−0.1892	−0.0033	−0.0033

Table 7
Empirical MSEs of HE, OLS and CE ($m = 10$, $n = 20$).

	True	HE	OLS	CE	True	HE	OLS	CE
β_1	0.5	0.0245	0.2851	0.0212	0.5	0.0250	0.2899	0.0229
β_2	1	0.0210	0.3248	0.0210	1	0.0211	0.3235	0.0207
β_3	1	0.0240	0.2357	0.0240	1	0.0267	0.2362	0.0265
β_4	1	0.0296	0.2151	0.02889	1	0.0306	0.2204	0.0313
β_5	1	0.0209	0.1608	0.0188	1	0.0226	0.1617	0.0221
γ_1	0	0.1264	0.3825	0.3825	0.2	0.1264	0.3914	0.3914
β_1	0.5	0.0247	0.2849	0.0270	0.5	0.0250	0.2842	0.0366
β_2	1	0.0208	0.3144	0.0219	1	0.0213	0.3232	0.0260
β_3	1	0.0359	0.2381	0.0344	1	0.0523	0.2395	0.0483
β_4	1	0.0427	0.2166	0.0510	1	0.0624	0.2174	0.0855
β_5	1	0.0275	0.1652	0.0374	1	0.0376	0.1619	0.0675
γ_1	0.5	0.1341	0.3900	0.3900	0.8	0.1445	0.3832	0.3832
β_1	0.5	0.0249	0.2862	0.0446	0.5	0.0250	0.2862	0.0552
β_2	1	0.0215	0.3229	0.0293	1	0.0217	0.3215	0.0335
β_3	1	0.0693	0.2351	0.0629	1	0.0886	0.2355	0.0794
β_4	1	0.0791	0.2160	0.1151	1	0.1053	0.2159	0.1589
β_5	1	0.0470	0.1689	0.0951	1	0.0575	0.1610	0.1280
γ_1	1	0.1490	0.3949	0.3949	1.2	0.1579	0.3800	0.3800

Table 8
Regression coefficient estimates and empirical means and standard derivations by the bootstrap method.

Coeff	HE	HE-BM(SD)	OLS	OLS-BM(SD)	CE	CE-BM(SD)
β_1	3.1275	3.1266(0.0990)	2.7718	2.7917(0.4556)	3.1261	3.1259(0.0989)
β_2	0.1124	0.1196(0.0922)	−0.1330	−0.1421(0.1486)	0.1220	0.1135(0.0856)
β_3	−0.1040	−0.0895(0.1239)	−0.5338	−0.4542(0.4568)	−0.1018	−0.8989(0.1223)
β_4	0.0918	0.0951(0.1188)	0.1148	0.1069(0.3015)	0.0922	0.0977(0.1169)
β_5	0.1572	0.1440(0.1137)	0.2034	0.1266(0.2412)	0.1588	0.1507(0.1121)
γ_1	0.0678	0.0566(0.1589)	0.2204	0.2223(0.2098)	0.2323	0.2223(0.2098)

HE-BM(SD): empirical mean (standard derivation) of HE by the bootstrap method.
OLS-BM(SD): empirical mean (standard derivation) of OLS by the bootstrap method.
CE-BM(SD): empirical mean (standard derivation) of CE by the bootstrap method.

Table 9
Regression coefficient estimates and empirical coefficient of variation by the bootstrap method.

Coeff	HE	HE-CV	OLS	OLS-CV	CE	CE-CV
β_1	3.1275	0.0316	2.7718	0.1632	3.1261	0.0465
β_2	0.1124	0.7709	−0.1330	−1.0457	0.1220	0.4900
β_5	0.1572	0.7895	0.2034	1.9052	0.1588	0.7439

HE-CV: empirical coefficient of variation of HE by the bootstrap method.
OLS-CV: empirical coefficient of variation of OLS by the bootstrap method.
CE-CV: empirical coefficient of variation of CE by the bootstrap method.

3.2. A real application

We apply the proposed method to the GPA data described in the Introduction Section. There are total 40 students in the data where 20 students are from the United States and 20 students are from other countries. We standardize the variables of masters GPA (response variable), undergraduate GPA (x_2), GRE verbal (x_3), GRE quantitative (x_4), GRE analytic (x_5), and TOEFL (z_1) by subtracting the empirical mean and dividing them by the empirical standard deviation. Table 8 shows the regression coefficient estimates using HE, OLS and CE. From Table 8, we find that the master GPA is positively correlated with undergraduate GPA, GRE quantitative, GRE analytic and TOEFL, while it is negatively correlated with GRE verbal. We also find that TOEFL is weakly associated with the master GPA by using HE. In order to demonstrate the robustness of our hybrid estimate, we repeat these three estimates 10,000 times by using the bootstrap procedure. Then we obtain the empirical means and standard deviations of three estimates. All results are displayed in Table 8. We can also see that the empirical standard derivation of HE is mostly smaller than those of OLS and CE.

Because the estimated values of β_3, β_4 and γ_1 are too small (as compared to the corresponding estimated standard errors). For β_1, β_2 and β_5, we use the coefficient of variation to explore the performance of HE. Table 9 shows the results for β_1, β_2 and β_5, which includes the regression coefficient estimates and the coefficient of variation by using the bootstrap procedure.

4. Discussion

Block missing data is commonly come across in practice. Imputation is often used when the missing part of the data has the same statistical distribution with the non-missing data, and also the missing mechanism is random. However, not all the missing data satisfy the above assumptions, such as the GPA data set with TOEFL. In this paper we propose a hybrid estimator that can handle the situation with block missing data. Under some reasonable conditions, our proposed hybrid estimates are asymptotically unbiased. But bias is not ignorable when the total size is small. When the missing rate is low, our proposed hybrid estimate (HE) overall performs better than the ordinary least square estimate only based on complete samples and combined estimate (Han and Li, 2011), according to MSE. When the missing rate becomes larger and the absolute value of covariates observations from partly missing samples is not large (e.g. be bounded), although the empirical bias of our HE is large, the MSE is much smaller than the other two estimates. As expected, our estimate may not perform well with the high missing rate. Fortunately, the hybrid estimator has a closed form and can be computed easily.

The performance of hybrid estimate depends on how good the prior is. If the prior data follows the same distribution as the current data, then increasing its size should result in better estimation of β, i.e., the bias will be smaller. If the prior data distribution is similar but not the same as that of the current data, the bias will become bigger as n increase.

If we assume the orthogonality of variables $\mathbf{X}$ *and* $\mathbf{Z}$ *instead of that* $\mathbf{X}_-^\tau/m = o(1)$ *as m and n both go to infinity in Theorem 1, the conclusions of Theorem 1 still hold. Since the bias may not be a big concern in practice, one easy way is to orthogonalize* $\mathbf{X}_+$ *and* $\mathbf{Z}_+$*, that is, we generate* $\mathbf{Z}_+^*$ *such that* $\mathbf{X}_+^\tau Z_+^* = 0$*. Based on this, the bias in the parameter estimates of* $\mathbf{X}$ *and* $\mathbf{Z}$ *will vanish.*

We will further study using some hybrid regularization approach where β is penalized toward $\tilde{\beta}$ (the estimate based on $\mathbf{Y}_-$ and X_-), and the degree of regularization is estimated using cross-validation.

Acknowledgments

Research of Zhengbang Li is partly supported by Central China Normal University Grant (No. CCNU13A05032). Research of Qizhai Li is partially supported by the National Science Foundation of China (No. 61134013, 11371353). Research of Bo Li is partly supported by National Natural Science Foundation of China (No. 11201165) and Central China Normal University Grant (No. CCNU09A01002). We greatly appreciate the insightful comments of associated editor and 2 reviewers.

Appendix

Proof of Theorem 1. Because the conclusion in scenario (1) can be obtained easily from the conclusion in scenario (2), so herein we only present the proof for scenario (2).

Denote $\eta = (\beta^\tau,\ \gamma^\tau)^\tau$, $\eta^* = \left((\beta^*)^\tau, (\gamma^*)^\tau\right)^\tau$, and $\overset{\circ}{\eta} = \left((\overset{\circ}{\beta})^\tau, (\overset{\circ}{\gamma})^\tau\right)^\tau$. Let $l^{*(p)}(\eta)$ and $l^{(p)}(\eta)(p = 1, 2)$ be the p-order partial derivatives of $l^*(\theta)$ and $l(\theta)$ according to η respectively. Denote $h(\beta) = \ln\pi(\beta)$ and define $h^{(p)}(\beta)(p = 1, 2)$ to be the p-order partial derivatives of $h(\beta)$ according to β. Denote by $\mathbf{Z}_-$ the missing observations for covariate Z, which is an $m \times r$ matrix.

(i) By the definition of η^*, $l^{*(1)}(\eta^*) = 0$. So we have

$$-l^{*(1)}(\overset{\circ}{\eta}) = l^{*(1)}(\eta^*) - l^{*(1)}(\overset{\circ}{\eta}) = l^{*(2)}(\check{\eta})(\eta^* - \overset{\circ}{\eta}),$$

where $\check{\eta} = (\check{\beta}^\tau, \check{\gamma}^\tau)^\tau$ is between η^* and $\overset{\circ}{\eta}$, that is, $\check{\eta}$ is in the neighborhood of $\overset{\circ}{\eta}$ with radius $\|\eta^* - \overset{\circ}{\eta}\|_2$. Based on some algebras, we have $\eta^* - \overset{\circ}{\eta} = \left\{-\frac{1}{m} l^{*(2)}(\check{\eta})\right\}^{-1} \left\{\frac{1}{m} l^{*(1)}(\overset{\circ}{\eta})\right\}$, and

$$\begin{aligned}
-\frac{1}{m} l^{*(2)}(\check{\eta}) &= -\frac{1}{m} l^{(2)}(\check{\eta}) - \frac{1}{m} h^{(2)}(\check{\beta}) \\
&= \frac{1}{m\sigma^2}\begin{pmatrix} \mathbf{X}_+^\tau \mathbf{X}_+ & X_+^\tau \mathbf{Z}_+ \\ \mathbf{Z}_+^\tau \mathbf{X}_+ & Z_+^\tau \mathbf{Z}_+ \end{pmatrix} + \frac{1}{m\sigma^2}\begin{pmatrix} \mathbf{X}_-^\tau \mathbf{X}_- & \mathbf{0}_{k\times r} \\ \mathbf{0}_{r\times k} & \mathbf{0}_{r\times r} \end{pmatrix} = \mathbf{W}/\sigma^2 + o(1), \text{ a.s.}
\end{aligned}$$

By the central limit theorem, we have

$$\begin{aligned}
&\frac{1}{m} l^{*(1)}(\overset{\circ}{\eta}) = \frac{1}{m} l^{(1)}(\overset{\circ}{\eta}) + \frac{1}{m} h^{(1)}(\overset{\circ}{\beta}) = \frac{1}{m\sigma^2}\begin{pmatrix} \mathbf{X}_+^\tau(\mathbf{Y}_+ - \mathbf{X}_+\overset{\circ}{\beta} - \mathbf{Z}_+\overset{\circ}{\gamma}) \\ \mathbf{Z}_+^\tau(\mathbf{Y}_+ - \mathbf{X}_+\overset{\circ}{\beta} - \mathbf{Z}_+\overset{\circ}{\gamma}) \end{pmatrix} \\
&-\frac{1}{m\sigma^2}\begin{pmatrix} \mathbf{X}_-^\tau \mathbf{X}_-(\overset{\circ}{\beta} - \tilde{\beta}) \\ \mathbf{0}_{r\times 1} \end{pmatrix} = \frac{1}{m\sigma^2}\begin{pmatrix} \mathbf{X}_+^\tau \boldsymbol{\xi} \\ \mathbf{Z}_+^\tau \boldsymbol{\xi} \end{pmatrix} + \frac{1}{m\sigma^2}\begin{pmatrix} \mathbf{X}_-^\tau(\mathbf{Z}_-\gamma + \boldsymbol{\xi}_\star) \\ \mathbf{0}_{r\times 1} \end{pmatrix} \\
&\qquad = \frac{1}{m\sigma^2}\begin{pmatrix} \mathbf{X}_+^\tau \boldsymbol{\xi} \\ \mathbf{Z}_+^\tau \boldsymbol{\xi} \end{pmatrix} + \frac{\mathbf{X}_-^\tau}{m\sigma^2}\begin{pmatrix} \mathbf{Z}_-\gamma + \boldsymbol{\xi}_\star \\ \mathbf{0}_{r\times 1} \end{pmatrix} = o(1), \text{ a.s.}
\end{aligned}$$

where $\xi = (\varepsilon_1, \ldots, \varepsilon_m)^\tau$, $\xi_\star = (\varepsilon_{m+1}, \ldots, \varepsilon_{m+n})^\tau$, and $\varepsilon_1, \ldots, \varepsilon_{m+n},\ i.i.d. \sim N(0, 1)$.

Hence,

$$\eta^* - \overset{\circ}{\eta} = \left\{-\frac{1}{m} l^{*(2)}(\check{\eta})\right\}^{-1} \left\{\frac{1}{m} l^{(1)}(\overset{\circ}{\eta}) + \frac{1}{m} h^{(1)}(\overset{\circ}{\beta})\right\} \xrightarrow{0}, \text{ a.s.}$$

or $\eta^* \to \overset{\circ}{\eta}$ (a.s.). By the definition of $(\sigma^2)^*$, we have $(\sigma^2) \to (\overset{\circ}{\sigma})^2$(a.s.)

(ii) From (i), we have $\theta^* \to \overset{\circ}{\theta}$ (a.s.). and $-\frac{1}{m} l^{*(2)}(\check{\eta}) = -\frac{1}{m} l^{(2)}(\check{\eta}) - \frac{1}{m} h^{(2)}(\check{\beta}) \to \mathbf{W}/(\overset{\circ}{\sigma})^2$(a.s.). Since $\frac{1}{\sqrt{m}} l^{(1)}(\overset{\circ}{\eta}) \xrightarrow{D} N\left(0, \mathbf{W}/(\overset{\circ}{\sigma})^2\right)$,

we have

$$\begin{aligned}
\sqrt{m}\left((\beta^*)^\tau - (\overset{\circ}{\beta})^\tau, (\gamma^*)^\tau - (\overset{\circ}{\gamma})^\tau\right)^\tau &= \left\{-\frac{1}{m} l^{*(2)}(\check{\eta})\right\}^{-1} \left\{\frac{1}{\sqrt{m}} l^{*(1)}(\overset{\circ}{\eta})\right\} \\
&= \left\{-\frac{1}{m} l^{*(2)}(\check{\eta})\right\}^{-1} \left\{\frac{1}{\sqrt{m}} l^{(1)}(\overset{\circ}{\eta}) + o(1)\right\} \xrightarrow{D} N\left(0, (\overset{\circ}{\sigma})^2 \mathbf{W}^{-1}\right)
\end{aligned}$$

by Slutzky' theorem.

References

Afifi, A.A., Elashoff, R.M., 1966. Missing observations in multivariate statistics. I. Review of the literature. J. Amer. Statist. Assoc. 61, 595–604.
Anderson, T.W., 1957. Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. J. Amer. Statist. Assoc. 52, 200–203.
Chen, H.Y., Little, R., 1999. A test of missing completely at random for generalised estimating equations with missing data. Biometrika 86, 1–13.
Dahiya, R.C., Korwar, R.M., 1980. Maximum likelihood estiamtes for a bivariate normal distribution with missing data. Ann. Statist. 8, 687–692.
Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. Ser. B Methodol. 39, 1–38.
Han, C.P., Li, Y., 2011. Regression anlaysis with block missing values and variable selection. Pakistan J. Statist. Oper. Res. 2, 391–400.
Hocking, R.R., Smith, W.B., 1968. Estimation of parameters in the multivariate normal distribution with missing observations. J. Amer. Statist. Assoc. 63, 159–173.
Lipsitz, S.R., Ibrahim, J.G., Fitzmaurice, G.M., 1999. Likelihood methods for incomplete longitudinal binary responses with incomplete categorical covariates. Biometrics 55, 214–223.
Little, R.J., Rubin, D.B., 2000. Statistical Analysis With Missing Data, seconnd ed. Wiley, New York.
Nicholson, G.J., 1957. Estimation of parameters from incomplete multivariate samples. J. Amer. Statist. Assoc. 52, 523–526.
Ratkowsky, D.A., 1974. Maximum likelihood estimation in small incomplete samples from the bivariate normal distribution. J. R. Stat. Soc. Ser. C. Appl. Stat. 23, 180–189.
Rochon, J., Helms, R.W., 1984. Maximum likelihood estimation for incomplete repeated-measures experiments under an ARMA covariance structure. Biometrics 45, 207–218.
Rubin, D.B., 1976. Inference and missing data. Biometrika 63, 581–592.
Rubin, D.B., 1987. Multiple Imputation for Nonresponse in Surveys. Wiley, New York.
Sexton, J., Laake, P., 2009. Stochastic approximation boosting for incomplete data problems. Biometrics 65, 1156–1163.
Sundberg, R., 1974. Maximum likelihood theory for incomplete data from an exponential family. Scand. J. Statist. 1, 49–58.
Titterington, D.M., 1984. Recursive parameter estimation using incomplete data. J. R. Stat. Soc. Ser. B 46, 257–267.
Woolson, R.F., Clarke, W.R., 1984. Analysis of categorical incomplete longitudinal data. J. R. Stat. Soc. Ser. A 147, 87–99.
Yuan, A., 2009. Bayesian frequentist hybrid inference. Ann. Statist. 37, 2083–2095.
Yuan, A., De Gooijer, J., 2013. Asymptotically informative prior for Bayesian analysis. Commun. Statist. Theory Methods (in press).