SSC Annual Meeting, June 2017
Proceedings of the Survey Methods Section

A UNIFIED EMPIRICAL LIKELIHOOD APPROACH TO TESTING MCAR AND SUBSEQUENT ESTIMATION

Shixiao Zhang[1], Peisong Han[2] and Changbao Wu[3]

ABSTRACT

For estimation with missing data, a crucial step is to determine if the data are missing completely at random (MCAR), in which case a complete-case analysis would suffice. Most existing tests for MCAR do not provide a method for subsequent estimation once the MCAR is rejected. In the setting of estimating the means of some response variables that are subject to missingness, we propose a unified approach to testing MCAR and the subsequent estimation. Upon rejecting MCAR, the same set of weights used for testing can then be used for estimation. The resulting estimators are consistent if the missingness of each response variable depends only on a set of fully observed auxiliary variables and the true outcome regression model is among the user-specified functions for deriving the weights. The proposed procedure is based on the calibration idea from survey sampling literature and the empirical likelihood theory. Simulation results show that the proposed strategy performs well for both testing and subsequent estimation.

KEY WORDS: Calibration, Empirical likelihood, Missing completely at random, Missingness mechanism

RÉSUMÉ

Pour une estimation avec des données manquantes, il est essentiel de déterminer si les données sont manquantes de façon complètement aléatoire (MFCA) et dans ce cas, une analyse de cas complets devrait suffire. La plupart des tests existants avec données MFCA ne fournissent aucune méthode d?estimation subséquente en cas de rejet de l?hypothèse MCFA. Pour l?estimation de la moyenne de certaines variables de réponse avec données manquantes, nous proposons une approche unifiée pour tester les données MCFA et l?estimation subséquente. S?il y a rejet de données MCFA, le même ensemble de poids utilisé pour le test peut aussi servir pour l?estimation. Les estimateurs qui en résultent sont convergents si l?état manquant de chaque variable de réponse dépend seulement d?un ensemble de variables auxiliaires complètement observées et que le véritable modèle de régression de chaque variable réponse est parmi les fonctions spécifiées par l?utilisateur pour dériver les poids. La procédure proposée se fonde sur le principe de calage tiré de la littérature de la théorie d'échantillonnage et de la théorie empirique de la vraisemblance. Les résultats de simulation indiquent que la stratégie proposée performe bien pour les tests et l'estimation subséquente

MOTS CLÉS : Calage, vraisemblance empirique, manquant complètement aléatoirement, mécanisme de non-réponse

1 INTRODUCTION

1.1 Description of the Problem

Data collected from statistical studies are often incomplete. There are three widely adopted missingness mechanisms in the missing-data literature (e.g., Little and Rubin 2002): missing completely at random (MCAR) where the missingness does not depend on either the observed or the missing data, missing at random (MAR) where the missingness depends on the observed but not the missing data, and missing not at random (MNAR) where the missingness depends on both the observed and the missing data. Most existing methods for missing-data analysis

[1]Department of Statistics and Actuarial Science, University of Waterloo, 200 University Ave. W., Waterloo, Ontario N2L 3G1, Canada; Corresponding author: Shixiao Zhang; Email: s398zhan@uwaterloo.ca.
[2]Department of Biostatistics, School of Public Health, University of Michigan, 1415 Washington Heights, Ann Arbor, Michigan 48109-2029, U.S.A.
[3]Department of Statistics and Actuarial Science, University of Waterloo, 200 University Ave. W., Waterloo, Ontario N2L 3G1, Canada.

are developed under the MAR mechanism, largely due to the mathematical triviality of MCAR and complexity of MNAR. However, in cases where the data are indeed MCAR, a simple complete-case analysis would suffice without turning to other possibly complicated methods. Therefore, a crucial first step for analysis with missing data is to determine if the missingness mechanism is MCAR.

The most widely used test for MCAR mechanism was due to Little (1988). Although it was proposed in the setting of multivariate normal data, the test is asymptotically valid regardless of the distribution of the data. Later a variety of literature is devoted to testing of MCAR, see Diggle (1989), Ridout (1991), Park and Davis (1993), Chen and Little (1999), Kim and Bentler (2002), Jamshidian and Jalal (2010), Li and Yu (2015). Despite the importance of determining the missingness mechanism, the ultimate task of data analysis is usually the subsequent estimation and inference. All the aforementioned works, however, treat the testing for MCAR as a stand-alone problem without providing a natural way for subsequent estimation once the MCAR mechanism is rejected. Our contribution in this paper is to propose a test for MCAR that also takes the subsequent estimation into account, so that an estimator of the quantity of interest with desirable properties is readily available once the MCAR is rejected. Our test does not impose any parametric assumptions on the underlying data distribution.

Our proposed unified procedure for testing and subsequent estimation is based on the calibration idea used in survey sampling literature (Deville and Särndal 1992; Wu and Sitter 2001) combined with the empirical likelihood method (Owen 1988, 2001; Qin and Lawless 1994). Under the MCAR mechanism, the complete cases are a random sample from the population, and thus the calibration weights assigned to the complete cases should be uniform with some random perturbation. Therefore, a significant deviation of the calibration weights from the uniform weights provides evidence against MCAR. The calibration weights are derived by matching the weighted average of certain user-specified functions of the fully-observed covariates based on the complete cases of the response to the unweighted average of those functions based on the whole sample. The functions may be certain moments of the covariates or the regression models of the response variables on the covariates. Upon rejecting MCAR, the calibration weights can be readily used to construct an estimator that is the weighted average of the observed values of the response variables, and these estimators are consistent if the missingness of each response variable depends only on the covariates and the corresponding correct regression model is among the user-specified functions used for calibration. Such an estimation approach agrees with the multiply robust estimation procedure in recent missing-data literature (e.g., Han and Wang 2013; Chan and Yam 2014; Han 2014; Han 2016a, 2016b).

2 THE PROPOSED METHOD

We first consider the simple scenario where the missingness only occurs to one variable, denoted by Y, and a vector of auxiliary variables $\boldsymbol{X}$ is fully observed. Let R denote the missingness indicator, S the set of complete cases and n_1 the number of complete cases. Under MCAR, S is a random sample from the population, and thus the sample mean of $\boldsymbol{X}$ based on the complete cases should be close to the sample mean based on the whole sample since both are consistent estimators of $E(\boldsymbol{X})$. In other words, if we assign positive weights w_i to the subjects in S so that $\sum_{i\in S} w_i \boldsymbol{X}_i = n^{-1}\sum_{j=1}^{n} \boldsymbol{X}_j$ and $\sum_{i\in S} w_i = 1$, then the w_i can be chosen to be close to the uniform weight $1/n_1$ where the deviation occurs only due to randomness. Therefore, a measure of the deviation from these w_i to $1/n_1$ provides an assessment of whether MCAR holds.

In practice, the ultimate goal is usually to estimate $E(Y)$ regardless of whether Y is MCAR. The estimation is often carried out by fitting a regression model for $E(Y \mid \boldsymbol{X})$ and then taking the sample mean of the fitted values over the whole sample. It is clear that the argument in the previous paragraph on using $\boldsymbol{X}$ to form constraints also applies to regression models viewed as functions of $\boldsymbol{X}$. Therefore, we consider the weights w_i satisfying the constraints

$$w_i > 0 \quad (i \in S), \quad \sum_{i\in S} w_i = 1, \quad \sum_{i\in S} w_i \boldsymbol{h}(\boldsymbol{X}_i; \hat{\boldsymbol{\theta}}) = \frac{1}{n}\sum_{j=1}^{n} \boldsymbol{h}(\boldsymbol{X}_j; \hat{\boldsymbol{\theta}}), \tag{1}$$

where $\boldsymbol{h}(\boldsymbol{X}; \boldsymbol{\theta})$ is a d-dimensional vector of user-specified functions of $\boldsymbol{X}$, possibly depending on some parameter $\boldsymbol{\theta}$ that is estimated by $\hat{\boldsymbol{\theta}}$. For example, $\boldsymbol{h}(\boldsymbol{X}; \boldsymbol{\theta})$ may include different moments of $\boldsymbol{X}$ and/or different regression models for $E(Y \mid \boldsymbol{X})$, and in the latter case $\boldsymbol{\theta}$ is the vector of all the regression parameters. Following the formulation of the empirical likelihood (EL) method (e.g., Qin and Lawless 1994), we consider the weights $\hat{w}_i$ that maximize $\prod_{i\in S} w_i$ subject to the constraints in (1). It turns out that, under MCAR, $\hat{w}_i$ are the weights we referred to in the previous paragraph that are close to the uniform weights $1/n_1$ where the deviation occurs only due to randomness.

The constraints in (1) are constructed based on the intuition that S is a random sample from the population under MCAR. A natural question then is whether these constraints are still compatible, or in other words whether there still exist w_i satisfying (1), when Y is not MCAR. The answer is affirmative. It can be easily shown that (e.g., Han and Wang 2013)

$$E\left(w(Y,\boldsymbol{X})\left[\boldsymbol{h}(\boldsymbol{X};\boldsymbol{\theta})-E\{\boldsymbol{h}(\boldsymbol{X};\boldsymbol{\theta})\}\right]\mid R=1\right)=\boldsymbol{0},$$

where $w(Y,\boldsymbol{X})=1/\mathbb{P}(R=1\mid Y,\boldsymbol{X})$. Then the constraints in (1) are simply the data version of the above moment equality, and thus are compatible even when Y is not MCAR.

From the EL theory (e.g., Qin and Lawless 1994), the $\hat{w}_i$ that maximize $\prod_{i\in S}w_i$ subject to (1) are given by

$$\hat{w}_i=\frac{1}{n_1}\frac{1}{1+\hat{\boldsymbol{\rho}}^{\mathrm{T}}\hat{\boldsymbol{g}}(\boldsymbol{X}_i;\hat{\boldsymbol{\theta}})}\qquad i\in S,$$

where $\hat{\boldsymbol{\rho}}$ is the Lagrange multiplier solving

$$\frac{1}{n_1}\sum_{i\in S}\frac{\hat{\boldsymbol{g}}(\boldsymbol{X}_i;\hat{\boldsymbol{\theta}})}{1+\hat{\boldsymbol{\rho}}^{\mathrm{T}}\hat{\boldsymbol{g}}(\boldsymbol{X}_i;\hat{\boldsymbol{\theta}})}=\boldsymbol{0} \tag{2}$$

and $\hat{\boldsymbol{g}}(\boldsymbol{X}_i;\hat{\boldsymbol{\theta}})=\boldsymbol{h}(\boldsymbol{X}_i;\hat{\boldsymbol{\theta}})-n^{-1}\sum_{j=1}^{n}\boldsymbol{h}(\boldsymbol{X}_j;\hat{\boldsymbol{\theta}})$. From the EL theory again, under MCAR, we have $\hat{\boldsymbol{\rho}}=O_p(n^{-1/2})$, which implies that the $\hat{w}_i$ are indeed equal to $1/n_1$ with a higher order perturbation. Now define

$$T=\frac{-2\sum\limits_{i\in S}\log(n_1\hat{w}_i)}{1-n_1/n}, \tag{3}$$

which is a measure of discrepancy between the $\hat{w}_i$ and $1/n_1$. The following result shows that T can be used to test for MCAR.

Theorem 1. *Under H_0: Y is MCAR, the test statistic T given in (3) has an asymptotic χ^2-distribution with degree of freedom d.*

When the MCAR is rejected, the $\hat{w}_i$ can be directly used to construct an estimator $\hat{\mu}=\sum_{i\in S}\hat{w}_iY_i$ for the quantity of interest $\mu_0=E(Y)$. Under MAR where the missingness of Y depends on $\boldsymbol{X}$ but not on Y, $\hat{\mu}$ is consistent for μ_0 if $\boldsymbol{h}(\boldsymbol{X};\boldsymbol{\theta})$ contains a correctly specified regression model for $E(Y|\boldsymbol{X})$. Therefore, the usage of the weights $\hat{w}_i$ is two-fold: it provides a test for MCAR and an estimator for μ_0, and thus makes our proposed method more attractive than existing ones.

Now we consider the case where $\boldsymbol{Y}_i=(Y_{1i},\cdots,Y_{pi})^{\mathrm{T}}$ is the p-dimensional data vector we intend to collect from subject i, $i=1,\cdots,n$ and each component of $\boldsymbol{Y}$ is subject to missingness but the auxiliary variables $\boldsymbol{X}$ are still fully observed. Let $\boldsymbol{R}_i=(R_{1i},\cdots,R_{pi})^{\mathrm{T}}$ be the vector of missingness indicators for $\boldsymbol{Y}_i$ such that $R_{ki}=1$ if Y_{ki} is observed and $R_{ki}=0$ otherwise, $k=1,\cdots,p$. Let $\pi_k\equiv\mathbb{P}(R_k=1)$ and assume that $\pi_k>0$ without loss of generality. Let S_k denote the set of subjects with Y_k observed and n_k the number of subjects in S_k, $k=1,\ldots,p$. To test if Y_k is MCAR, we can directly apply the test statistic given in (3) to Y_k based on a d_k-dimensional vector of user-specified functions $\boldsymbol{h}_k(\boldsymbol{X};\boldsymbol{\theta}_k)$. Let $\hat{w}_{ki}$, $i\in S_k$, denote the resulting weights for the subjects in S_k. It follows from Theorem 1 that the test statistic

$$T_k=\frac{-2\sum\limits_{i\in S_k}\log(n_k\hat{w}_{ki})}{1-n_k/n}$$

asymptotically follows the χ^2-distribution with d_k degrees of freedom if Y_k is MCAR. Furthermore, using the T_k, we are able to construct a test statistic to test if $\boldsymbol{Y}$ is MCAR as shown in the following result.

Theorem 2. *Under H_0: $\boldsymbol{Y}$ is MCAR, the test statistic $T_{sum}=\sum\limits_{k=1}^{p}T_k$ asymptotically has the same distribution as $\sum_{l=1}^{m}\lambda_lQ_l$ where $m=d_1+\cdots+d_p$ and for $l=1,\ldots,m$, the Q_l are independent χ^2-distributed random variables with 1 degree of freedom and the λ_l are the eigenvalues of*

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{I}_{d_1} & \boldsymbol{\Sigma}_{12} & \cdots & \boldsymbol{\Sigma}_{1p} \\ \boldsymbol{\Sigma}_{12} & \boldsymbol{I}_{d_2} & & \vdots \\ \vdots & & \ddots & \\ \boldsymbol{\Sigma}_{1p} & \cdots & & \boldsymbol{I}_{d_p} \end{pmatrix}.$$

Here $\boldsymbol{I}_{d_k}$ is the identity matrix with dimension d_k and, for $k, r = 1, \ldots, p$ and $k \neq r$,

$$\begin{aligned} \boldsymbol{\Sigma}_{kr} &= \{\pi_k \pi_r (1-\pi_k)(1-\pi_r)\}^{-1/2} (\pi_{kr} - \pi_k \pi_r) \\ &\quad \times \left[E\{\boldsymbol{g}_k(\boldsymbol{\theta}_{k*})\boldsymbol{g}_k(\boldsymbol{\theta}_{k*})^T\}\right]^{-1/2} \left[E\{\boldsymbol{g}_k(\boldsymbol{\theta}_{k*})\boldsymbol{g}_r(\boldsymbol{\theta}_{r*})^T\}\right] \left[E\{\boldsymbol{g}_r(\boldsymbol{\theta}_{r*})\boldsymbol{g}_r(\boldsymbol{\theta}_{r*})^T\}\right]^{-1/2}, \end{aligned}$$

$\pi_k = \mathbb{P}(R_k = 1)$, $\pi_{kr} = \mathbb{P}(R_k = 1, R_r = 1)$ *and* $\boldsymbol{g}_k(\boldsymbol{\theta}_k) \equiv \boldsymbol{g}_k(\boldsymbol{X}; \boldsymbol{\theta}_k) = \boldsymbol{h}_k(\boldsymbol{X}; \boldsymbol{\theta}_k) - E\{\boldsymbol{h}_k(\boldsymbol{X}; \boldsymbol{\theta}_k)\}$.

The eigenvalues λ_l are not necessarily distinct (e.g., Imhof 1961). In practice, in order to determine the critical value for the asymptotic distribution of T_{sum}, $\boldsymbol{\Sigma}_{kr}$ can be consistently estimated by replacing π_{kr} and π_k with n_{kr}/n and n_k/n where n_{kr} is the number of subjects with Y_k and Y_r observed simultaneously, and the expectations can be estimated by sample averages. When the MCAR is rejected, the weights $\hat{w}_{ki}$ used for testing can then be used to construct an estimator for $E(Y_k)$: $\sum_{i=1}^n R_{ki}\hat{w}_{ki}Y_{ki}$. Following the same argument as before, such an estimator is consistent if the missingness of Y_k depends only on $\boldsymbol{X}$ and one component of $\boldsymbol{h}_k(\boldsymbol{X}; \boldsymbol{\theta}_k)$ is the correctly specified regression model for $E(Y_k \mid \boldsymbol{X})$.

The construction of constraints in (1) is flexible in the sense that, in principle, any user-specified functions of $\boldsymbol{X}$ can be considered. The use of moments of $\boldsymbol{X}$ is standard in survey sampling literature on the calibration method (e.g., Deville and Särndal 1992; Chen and Sitter 1999; Kim 2000). The use of regression models has become popular in recent literature on calibration-based missing data analysis (e.g., Wu and Sitter 2001; Han and Wang 2013; Chan and Yam 2014; Han 2014, 2016a, 2016b). Our extensive simulation study shows that, using moments of $\boldsymbol{X}$ tends to lead to more power for the proposed test compared to using regression models only. This makes intuitive sense because (1) holds for any functions of $\boldsymbol{X}$ whereas a regression model only represents a particular function. On the other hand, including a correctly specified regression model helps to achieve estimation consistency, as argued before in this section. Therefore, in practice we would recommend using both moments of $\boldsymbol{X}$ and regression models to construct the constraints in (1).

The power of the proposed test is also affected by the missingness mechanism of each Y_k. If the missingness mechanism does not depend on $\boldsymbol{X}$, then the proposed test has no power detecting deviation from MCAR because the constraints in (1) are all functions of $\boldsymbol{X}$. In addition, for estimation, the proposed procedure implicitly assumes a regression model of $\boldsymbol{Y}$ on $\boldsymbol{X}$. When this assumption is violated, the proposed weighted estimator will no longer be consistent.

3 SIMULATION STUDY

We use a simulation setup mimicing the one in Chen and Little (1999) to study the type I error of the proposed test under MCAR and the power under different missingness mechanisms. Three covariates are independently generated as $X_1 \sim \text{Uniform}(-1, 1)$, $X_2 \sim N(0, 1)$ and $X_3 \sim \text{Bernoulli}(0.5)$. Given the covariates, $\tilde{Y}_1$ and $\tilde{Y}_2$ are independently generated from $N(X_1 + 2X_2 + 3X_3, 1)$. The two response variables are then generated as $Y_1 = \tilde{Y}_1$ and $Y_2 = U\tilde{Y}_1 + (1-U)\tilde{Y}_2$ where $U \sim \text{Bernoulli}\{(1+X_1)/2\}$.

We follow steps similar to those in Chen and Little (1999) to create missing values. First, each subject is classified into one of two sets with probabilities p^s and $1 - p^s$, respectively. Then, in the first set, Y_2 is fully observed while Y_1 is missing with probability p_1^s; in the second set, Y_1 is fully observed while Y_2 is missing with probability p_2^s. The dependence of p^s, p_1^s and p_2^s on $\boldsymbol{X}$ and/or $\boldsymbol{Y}$ determines the missingness mechanism. Table 1 gives a list of some specific combinations of (p^s, p_1^s, p_2^s) we use in the simulation study, where the parameters α_1 and α_2 take different values corresponding to different degrees of departure from MCAR ($\alpha_1 = 0$ and $\alpha_2 = 0$). The missingness mechanism that each specific combination corresponds to is also given. To distinguish different combinations and make them easier to be referred to in Tables 2 and 3, each specific combination, except the one corresponding to MCAR, is assigned a code in the form of "letter-number", where "a" and "b" correspond to $p^s = 0.5$ and $p^s = (1 + X_1)/2$ and "1", "2" and "3" correspond to MAR with missingness depending only on $\boldsymbol{X}$, MAR with missingness depending on the observed response and MNAR, respectively.

Since the correct regression models for $E(Y_1|\boldsymbol{X})$ and $E(Y_2|\boldsymbol{X})$ are linear models with regressors X_1, X_2 and X_3, including both the first moment of $\boldsymbol{X}$ and those linear regression models in $\boldsymbol{h}(\boldsymbol{X};\boldsymbol{\theta})$ results in collinearity. Therefore, we simply take $\boldsymbol{h}(\boldsymbol{X};\boldsymbol{\theta}) = \boldsymbol{X}$. We compare the proposed test with the ones in Little (1988) and Chen and Little (1999). Simulation results are summarized based on 1000 replications with sample size $n = 200$ for each replication, and the significance level is set at 5%.

Table 2 contains results on the type I error under MCAR and the power under different missingness mechanisms. The overall performance of the proposed test is quite close to that of Little (1988), and both are better than the test of Chen and Little (1999). As pointed out by Chen and Little (1999), their test actually tests the unbiasedness of a set of generalized estimating equations rather than the MCAR mechanism, and thus the performance depends on the specific form of the estimating equations and does not always agree with the theoretical behaviour of a test for MCAR.

Table 3 shows the performance of the weighted estimators of $E(Y_1)$ and $E(Y_2)$ based on the calibration weights that were used to construct the test statistic. Under MCAR, both the proposed estimator $\hat{\mu}_k$ and the complete-case average estimator $\hat{\mu}_{k\text{cc}}$ have negligible bias, $k = 1, 2$. The estimator $\hat{\mu}_{k\text{cc}}$ loses its consistency when the missingness mechanism is no longer MCAR, demonstrated by its non-negligible relative bias in those cases. On the contrary, the proposed estimator $\hat{\mu}_k$ is still consistent in cases a-1 and b-1 where the missingness depends only on the fully observed covariates. Surprisingly, for the other cases a-2, a-3, b-2 and b-3, although $\hat{\mu}_k$ is theoretically not consistent, its relative bias is very small compared to that of $\hat{\mu}_{k\text{cc}}$. This observation that the calibration-based estimators have relatively small bias even if its theoretical consistency cannot be formally shown has also been noted in Han (2014, 2016a) and demonstrates the superiority of these estimators.

4 CONCLUSIONS

Ascertaining the missingness mechanism is always a crucial step in missing data analysis. While MAR is in general not testable, MCAR is, and under MCAR data analysis becomes fairly easy since a complete case analysis would be sufficient. We have proposed a nonparametric approach based on the empirical likelihood method to test MCAR. The proposed approach not only provides an alternative to existing tests, but more importantly, for the commonly seen scenarios with the presence of fully observed covariates, it leads to a unified procedure for estimation after the MCAR is rejected with little extra effort beyond the calculation of the test statistic. Existing tests, on the contrary, only focus on testing, and the estimation after MCAR is rejected has to invoke possibly completely different procedures.

REFERENCES

Chan, K. C. G. and Yam, S. C. P. (2014). Oracle, multiple robust and multipurpose calibration in a missing response problem. *Statistical Science*, **29(3)**, 380–396.

Chen, H. Y. and Little, R. J. A. (1999). A test of missing completely at random for generalized estimating equations with missing data. *Biometrika*, **86**, 1–13.

Deville, J. and Särndal, C. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, **87(418)**, 376–382.

Diggle, P. J. (1989). Testing for random dropouts in repeated measurement data. *Biometrics*, **45(4)**, 1255–1258.

Han, P. (2014). Multiply robust estimation in regression analysis with missing data. *Journal of the American Statistical Association*, **109(507)**, 1159–1173.

Han, P. (2016a). Combining inverse probability weighting and multiple imputation to improve robustness of estimation. *Scandinavian Journal of Statistics*, **43**, 246–260.

Han, P. (2016b). Intrinsic efficiency and multiple robustness in longitudinal studies with drop-out. *Biometrika*, **103(3)**, 683–700.

Han, P. and Wang, L. (2013). Estimation with missing data: beyond double robustness. *Biometrika*, **100(2)**, 417–430.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, **47(260)**, 663–685.

Imhof, J. P. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika*, **48(3-4)**, 419–426.

Jamshidian, M. and Jalal, S. (2010). Tests of homoscedasticity, normality and missing completely at random for incomplete multivariate data. *Psychometrika*, **75(4)**, 649–674.

Kim, K. H. and Bentler, P. M. (2002). Tests of homogeneity of means and covariance matrices for multivariate incomplete data. *Psychometrika*, **67(4)**, 609–624.

Li, J. and Yu, Y. (2015). A nonparametric test of missing completely at random for incomplete multivariate data. *Psychometrika*, **80(3)**, 707–726.

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, **83(404)**, 1198–1202.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. John Wiley & Sons, Inc. New York, 2 edition.

Owen, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75(2)**, 237–249.

Owen, A. (2001). *Empirical likelihood*. Chapman & Hall/CRC Press, New York.

Park, T. and Davis, C. S. (1993). A test of the missing data mechanism for repeated categorical data. *Biometrics*, **49(2)**, 631–638.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Annals of Statistics*, **22(1)**, 300–325.

Ridout, M. S. (1991). Testing for random dropouts in repeated measurement data (reader reaction). *Biometrics*, **47(4)**, 1619–1621.

Wu, C. and Lu, W. (2016). Calibration weighting methods for complex surveys. *International Statistical Review*, **84(1)**, 79–98.

Wu, C. and Sitter, R. R. (2001). A model-calibration approach to using complete auxiliary information from survey data. *Journal of the American Statistical Association*, **96**, 185–193.

Table 1: The combinations of (p^s, p_1^s, p_2^s) used in Simulation Study

p^s	p_1^s	p_2^s	Mechanism	code
0.5	$\{1+\exp(0.5)\}^{-1}$	$\{1+\exp(0.5)\}^{-1}$	MCAR	
0.5	$\{1+\exp(0.5-\alpha_1/2+\alpha_1 X_2)\}^{-1}$	$\{1+\exp(0.5-\alpha_2/2+\alpha_2 X_2)\}^{-1}$	MAR	a-1
0.5	$\{1+\exp(0.5-\alpha_1/2+\alpha_1 Y_2)\}^{-1}$	$\{1+\exp(0.5-\alpha_2/2+\alpha_2 Y_1)\}^{-1}$	MAR	a-2
0.5	$\{1+\exp(0.5-\alpha_1/2+\alpha_1 Y_1)\}^{-1}$	$\{1+\exp(0.5-\alpha_2/2+\alpha_2 Y_2)\}^{-1}$	MNAR	a-3
$(1+X_1)/2$	$\{1+\exp(0.5-\alpha_1/2+\alpha_1 X_2)\}^{-1}$	$\{1+\exp(0.5-\alpha_2/2+\alpha_2 X_2)\}^{-1}$	MAR	b-1
$(1+X_1)/2$	$\{1+\exp(0.5-\alpha_1/2+\alpha_1 Y_2)\}^{-1}$	$\{1+\exp(0.5-\alpha_2/2+\alpha_2 Y_1)\}^{-1}$	MAR	b-2
$(1+X_1)/2$	$\{1+\exp(0.5-\alpha_1/2+\alpha_1 Y_1)\}^{-1}$	$\{1+\exp(0.5-\alpha_2/2+\alpha_2 Y_2)\}^{-1}$	MNAR	b-3

Table 2: Results on Type I error under MCAR and power under different missingness mechanisms for Simulation Study based on $n = 200$ and 1000 replications. The significance level is set to be 5%. The numbers are percentages.

		Little	C&L	T_{sum}	Little	C&L	T_{sum}
α_1	α_2	(a) $p^s = 0.5$			(b) $p^s = (1+X_1)/2$		
		MCAR					
0	0	4	17.9	5	–	–	–
		a-1 MAR			b-1 MAR		
0.3	-0.3	13.7	19.4	20.3	99.6	17.5	100
0.6	-0.3	32.2	17.8	42.6	99.9	16.7	100
0.3	0.3	26.3	19.4	21	99.6	16.7	100
0.6	0.3	53.4	16.9	46.4	100	15.3	100
		a-2 MAR			b-2 MAR		
0.3	-0.3	83.5	25.3	86.3	100	28.2	100
0.6	-0.3	99.1	34.6	99	100	38.2	100
0.3	0.3	96.7	31.7	92.6	100	31.3	100
0.6	0.3	100	41.6	99.9	100	39.7	100
		a-3 MNAR			b-3 MNAR		
0.3	-0.3	72.9	21.5	85.7	100	21.6	100
0.6	-0.3	97	25.6	99	100	22.7	100
0.3	0.3	95.9	27.6	93.3	100	24.6	100
0.6	0.3	100	30.5	100	100	25.8	100

Little: the test in Little (1988). C&L: the test in Chen and Little (1999). T_{sum}: our proposed test.

Table 3: Results on estimation of $E(Y_1) = E(Y_2) = 1.5$ using the calibration weights for Simulation Study based on $n = 200$ and 1000 replications. The numbers have been multiplied by 100.

		Estimation of $E(Y_1)$				Estimation of $E(Y_2)$			
		$\hat{\mu}_1$		$\hat{\mu}_{1cc}$		$\hat{\mu}_2$		$\hat{\mu}_{2cc}$	
α_1	α_2	rBias	RMSE	rBias	RMSE	rBias	RMSE	rBias	RMSE
MCAR									
0	0	0	20	1	22	0	20	0	21
a-1 MAR									
0.3	-0.3	0	20	6	24	0	20	-5	23
0.6	0.3	0	20	12	29	0	20	7	24
a-2 MAR									
0.3	-0.3	2	20	17	33	-1	20	-20	37
0.6	0.3	2	20	26	44	2	20	17	33
a-3 MNAR									
0.3	-0.3	3	21	18	34	-3	20	-21	39
0.6	0.3	4	21	28	46	3	20	18	35
b-1 MAR									
0.3	-0.3	1	20	12	29	0	20	-10	26
0.6	0.3	1	20	19	36	0	20	1	22
b-2 MAR									
0.3	-0.3	1	20	22	39	-2	20	-27	46
0.6	0.3	2	20	31	51	2	20	13	28
b-3 MNAR									
0.3	-0.3	3	21	24	41	-3	20	-28	48
0.6	0.3	5	21	34	55	3	20	13	29

$\hat{\mu}_k$ and $\hat{\mu}_{kcc}$: estimators of $E(Y_k)$ based on our proposed procedure and based on complete-case analysis, respectively, $k = 1, 2$. rBias: relative bias $1000^{-1} \sum_{b=1}^{1000} \{\hat{\mu}_{kb} - E(Y_k)\}/E(Y_k)$, where $\hat{\mu}_{kb}$ is the estimate of $E(Y_k)$ from the bth replication. RMSE: root mean square error.