



Empirical likelihood inference in general linear model with missing values in response and covariates by MNAR mechanism

Fayyaz Bahari¹ · Safar Parsi¹ · Mojtaba Ganjali²

Received: 22 October 2017 / Revised: 7 January 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

In this paper, we utilize a general linear model for analyzing data with missing values in some covariates and response variable. Our aim is to fit a general linear model and to construct a confidence region for the parameters of the general linear model based on the empirical likelihood ratio function. Also, we assume that missing data may happen in covariates or in response variable or in both of them with missing not at random mechanism where the probability of missing a datum is specified by a logistic model. We use inverse probability weights and an augmented method as the auxiliary condition of empirical likelihood to estimate parameters of the general linear model. Asymptotic properties of the empirical log-likelihood ratio are investigated whether the exponential tilting parameter is known or estimated by the follow-up sample. The asymptotic normality of estimators is also proved. Some simulation studies are used to illustrate the performance of our model for different sample sizes. Also, a real dataset is studied by the proposed methods.

Keyword General linear model · Missing data · Exponential tilting · Augmented method · Inverse probability weights method · Empirical log-likelihood ratio

✉ Safar Parsi
parsi@uma.ac.ir

Fayyaz Bahari
fayyaz.bahari@yahoo.com

Mojtaba Ganjali
M-Ganjali@sbu.ac.ir

¹ Department of Statistics, Faculty of Mathematical Sciences, University of Mohaghegh Ardabil, Ardabil, Iran

² Department of Statistics, Faculty of Mathematical Sciences, Shahid Beheshti University, Tehran, Iran

1 Introduction

In most regression studies, the relationship between the response variable and the covariates is not linear and fitting a linear model is not appropriate. To solve this problem, we use a general version of the linear model, the general linear model. A general linear model represents the relationship between a response Y and a vector function of covariates X with dimension m that can be formulated as follow:

$$Y = h^T(X)\beta + \epsilon, \quad (1)$$

where $h(\cdot)$ is a known vector function of dimension d , T is used for transposing, β is an unknown parameter vector of dimension d and ϵ is the sampling error with mean zero and constant variance. Also, it is assumed that the conditional expectation of ϵ follows the conditions $E(\epsilon|X = x) = 0$ and $E(\epsilon^2|X = x) < \infty$.

In this paper, we assume that missing data can be occurred both in covariates and in response variables but it is necessary to have at least one variable which is fully observed because to estimate unknown functions based on kernel method, it is necessary to have at least one variable which is fully observed (at least response variable or one of the covariate variables). It is assumed that V contains all the variables, Z contains fully observed variables and U contains incomplete variables. Therefore, the data for the i th individual will be as follow:

$$(x_i, y_i, \delta_i) = (v_i, \delta_i) = (z_i, u_i, \delta_i), \quad i = 1, \dots, n,$$

where $\delta_i = 1$ if the i th individual of variable U is observed and otherwise $\delta_i = 0$.

By Rubin's classification (1976), we say missing mechanism is MCAR if the probability of observing a datum does not depend on the data, missing mechanism is MAR if the probability of observing a datum does not depend on the missing data and missing mechanism is MNAR if the probability of observing a datum depends on the missing and observed data. Most researchers consider missing at random (MAR) mechanism as the common mechanism for statistical inferences with missing data. However, missing mechanism depends on missed data in many cases such as survey data, longitudinal data, clinical data etc., and using missing not at random mechanism can be more reasonable. Moreover, MNAR mechanism is more general than MAR mechanism and MAR mechanism can be regarded as a submodel of MNAR mechanism. Throughout this article, we assume the MNAR mechanism as a mechanism of the missing data.

In MNAR mechanism, missingness probability is unknown and moreover, the probability of observing or missing an individual depends on the variable that contains missing data, for these reasons estimating of missingness probability is more complex than that of MAR mechanism. For this reason, Kim and Yu (2011) determined a semi-parametric method to estimate mean function based on the exponential tilting model which is considered for the missingness probability. Zhao et al. (2013) improved the Kim and Yu (2011) method by using more auxiliary information in the empirical likelihood ratio function. Furthermore, Tang and Zhao (2013) used the empirical log-likelihood ratio to inference about the nonlinear regression models. On the other hand,

Niu et al. (2014) used this empirical method for fitting a linear model. They showed that the log-likelihood ratio is asymptotically convergent to a weighted chi-square distribution and $\sqrt{n}$ -consistent estimators of parameters are asymptotically convergent to the normal distribution. In this way, they have constructed the confidence intervals based on the empirical log-likelihood ratio which is introduced by Owen (1988, 1990). Also, the improved version of this method is given by Zhao et al. (2015) to inference about the regression model. Since in the empirical log-likelihood method, the confidence interval does not require prior assumptions and pivotal, it has a widespread use in statistical inferences. Also, in Owen's mentioned method, confidence regions are constructed based on an appropriate region that is dominated by the empirical log-likelihood ratio. In addition, Niu et al. (2014) have considered a simple linear model where missing data happens by MNAR mechanism in the response variable and they have considered some imputation methods to deal with missing data problems.

In this paper, we consider a general linear model where missing data can occur in the covariates and response variables by an MNAR mechanism. We use three auxiliary conditions of estimating equations in empirical likelihood functions to inference about the general linear model. In this way, we apply the similar methods used in Creemers et al. (2011) to add partly observed data in inferences. Also, we estimate exponential tilting parameter from a follow-up sample. In this paper, we have assumed that missing data follows from a logistic model. To see inferences without this assumption see also Genback et al. (2015) where they inference about a regression model with missing data by a nonignorable mechanism.

The rest of this paper is organized as follow. In Sect. 2, we introduce three methods to estimate parameters of the general linear model and we will illustrate the empirical details of this three methods. In Sect. 3, we will present asymptotic properties of the three methods. In Sect. 4, some simulation studies are used to see the performance of the different methods under the different missing mechanism schemes and different sample sizes assumptions. In Sect. 6, the details of some points are discussed. Also, the details and proofs of the theorems are given in the Appendix.

2 Methods

In this section, we propose three methods for analyzing general linear model with missing data by MNAR mechanism where missing data occurs in covariates or response variable. To achieve this aim, it is enough to have at least one variable without missing data. Our aim is to estimate parameters of the general linear model and then constructing the appropriate confidence intervals for the parameters of this model.

Let $\pi(v_i)$ be the probability of observing the i th datum. Therefore, $\pi(v_i)$ takes the following form under the MNAR assumption:

$$\pi(v_i) = Pr(\delta_i = 1|x_i, y_i) = Pr(\delta_i = 1|z_i, u_i) = Pr(\delta_i = 1|v_i), \quad i = 1, \dots, n. \quad (2)$$

In real data analyses, $\pi(v_i)$ is unknown and we have to estimate it. It seems that considering a model for missingness pattern is necessary in MNAR case. As Kim and

Yu (2011) done, we consider the following logistic model for missing data probability:

$$\pi(v_i) = \Pr(\delta_i = 1|z_i, u_i) = \frac{e^{g(z_i) + \phi c(u_i)}}{1 + e^{g(z_i) + \phi c(u_i)}}, \quad i = 1, \dots, n, \quad (3)$$

where $g(\cdot)$ is an unknown function, ϕ is an unknown parameter and $c(\cdot)$ is a real-valued function of variables with missing data. The conditional odds of δ_i can be written as follow:

$$O(z_i, u_i) = \frac{\Pr(\delta_i = 0|z_i, u_i)}{\Pr(\delta_i = 1|z_i, u_i)} = \frac{1}{\pi(v_i)} - 1 = e^{-g(z_i) + \gamma c(u_i)}, \quad i = 1, \dots, n. \quad (4)$$

where $\gamma = -\phi$. On the other hand, based on the exponential tilting model which is mentioned on Kim and Yu (2011), we can write:

$$f_0(u_i|z_i) = \frac{e^{\gamma c(u_i)}}{E(e^{\gamma c(U_i)}|z_i, \delta_i = 1)} f_1(u_i|z_i), \quad i = 1, \dots, n, \quad (5)$$

where for $\delta = 0$, $f_0(u|z)$ is the conditional density of U given Z and similarly, $f_1(u|z)$ is the conditional density of U given Z for $\delta = 1$. Also for simplicity, we define $\alpha(z_i, \gamma) = \frac{O(z_i, u_i)}{e^{\gamma c(u_i)}}$. Under the logistic model of Eq. (3) and exponential tilting model of Eq. (5), by replacing $O(Z_i, U_i)$ with $\frac{1}{\pi(V_i)} - 1$ and some mathematical operations, it is easy to prove that:

$$\begin{aligned} E(\delta_i O(Z_i, U_i)|Z_i = z_i) &= E\left(\frac{\delta_i}{\pi(V_i)} - \delta_i|Z_i = z_i\right) \\ &= 1 - E(\delta_i|Z_i = z_i) \\ &= E(1 - \delta_i|Z_i = z_i), \end{aligned} \quad (6)$$

in addition, by replacing $O(Z_i, U_i)$ with $\alpha(Z_i, \gamma)e^{\gamma c(u_i)}$ and some mathematical operations, we can conclude that:

$$E(\delta_i O(Z_i, U_i)|Z_i = z_i) = \alpha(Z_i, \gamma)E(\delta_i e^{\gamma c(u_i)}|Z_i = z_i). \quad (7)$$

Therefore, by combining Eqs. (6) and (7), we can conclude,

$$\begin{aligned} \alpha(Z_i, \gamma)E(\delta_i e^{\gamma c(u_i)}|Z_i = z_i) &= E(1 - \delta_i|Z_i = z_i) \\ &= E(\delta_i O(Z_i, U_i)|Z_i = z_i). \end{aligned} \quad (8)$$

Thus, based on Eq. (8), an empirical estimator of $\alpha(z_i, \gamma)$ will be as:

$$\hat{\alpha}(z_i, \gamma) = \frac{\sum_{j=1}^n (1 - \delta_j) K_h(z_j - z_i)}{\sum_{j=1}^n \delta_j e^{\gamma c(u_j)} K_h(z_j - z_i)}, \quad (9)$$

where $K_h(\cdot)$ is a kernel function with smoothing parameter h . Also in a similar way, an empirical estimator of $\pi(v_i)$ based on $\hat{\alpha}(z_i, \gamma)$ can be obtained as:

$$\hat{\pi}(v_i) = \{1 + \hat{\alpha}(z_i, \gamma)e^{\gamma c(u_i)}\}^{-1}. \quad (10)$$

Above estimators are valid when γ is known. It is considered known in Rotnitzky et al. (1998) to access the sensitivity of the analysis. But, in most studies it is unknown and we have to estimate it. An estimation of tilting parameter γ can be found parametrically by an estimating equation based on a follow up sample. For more details about validation sample and estimation of tilting parameter see also Kim and Yu (2011). The estimating equations to find γ is given in the next section.

2.1 Empirical likelihood inference based on the inverse probability weights

We assume that auxiliary random vectors based on the inverse probability weights to be as follow:

$$\psi_{i,IPW}(\beta) = \frac{\delta_i}{\pi(v_i)} S(v_i, \beta), \quad i = 1, \dots, n, \quad (11)$$

where $S(v_i, \beta) = S(y_i, x_i, \beta) = h(x_i)(y_i - h^T(x_i)\beta)$, is the score function which is deduced based on the ordinary least square method and it can be obtained from the other estimating methods such as maximum likelihood method. By the above definition, we can formulate the empirical likelihood model in a simple way. It is easy to prove that $E(\psi_{i,IPW}(\beta)) = 0$. Therefore, the inverse probability weighted empirical likelihood (IPWEL) function can be defined as follow:

$$L_{IPW}(\beta) = \sup \left\{ \prod_{i=1}^n n p_i \mid p_i > 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \psi_{i,IPW}(\beta) = 0 \right\}, \quad (12)$$

where p_i is the probability mass of the i th member of the sample and as mentioned in the above equation $\sum_{i=1}^n p_i = 1$. Without considering auxiliary condition, $\sum_{i=1}^n p_i \psi_{i,IPW}(\beta) = 0$, in the above IPWEL function, we can estimate $\hat{p}_i = \frac{1}{n}$. The inverse probability weighted log-likelihood ratio (IPWLR) function for parameter β with some corrections (minus twice of the empirical log-likelihood ratio to obtain its asymptotic distribution) will be as follow:

$$l_{IPW}(\beta) = -2 \sup \left\{ \sum_{i=1}^n \log(n p_i) \mid p_i > 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \psi_{i,IPW}(\beta) = 0 \right\}. \quad (13)$$

Therefore, by Lagrange method, one can obtain that, $\hat{p}_i = \frac{1}{n(1 + \lambda^T \psi_{i,IPW}(\beta))}$, where λ is a Lagrange multiplier real valued vector with dimension d . The IPWLR function will be as follow:

$$l_{IPW}(\beta) = 2 \sum_{i=1}^n \log(1 + \lambda^T \psi_{i,IPW}(\beta)). \quad (14)$$

The empirical maximum likelihood estimator of β by IPW method can be calculated by maximizing $-l_{IPW}(\beta)$ over all domain of β and λ . Based on the IPW method, we estimate β_{IPW} with Lagrange multiplier method by using the following estimating equations:

$$\begin{cases} T_{1,IPW} = \frac{1}{n} \sum_{i=1}^n \frac{\psi_{i,IPW}(\beta)}{1 + \lambda^T \psi_{i,IPW}(\beta)} = 0 \\ T_{2,IPW} = \frac{1}{n} \sum_{i=1}^n \frac{\lambda \delta_i h(x_i) h^T(x_i)(\beta)}{\pi(v_i)(1 + \lambda^T \psi_{i,IPW}(\beta))} = 0. \end{cases} \quad (15)$$

2.2 Empirical likelihood inference based on the augmented method

In IPWEL inferences, we have ignored the partly missed data where is partly observed too. This may decrease the efficiency of our estimators. for example, the big missingness probability of a datum may depend on the magnitude of the data. In this case, we may loss the data with large values or with small values. On the other hand, we loss partly observed data that is useful in our inferences and may increase the variance of the estimator. To see more details about missing data effect on efficiency of estimators see also Carpenter et al. (2006). Therefore, we consider two other methods where these methods can use the informations of the partly observed data. We also use an augmented method that augments expectation of auxiliary mean score vector by an appropriate weight to the inverse probability weight auxiliary condition as follow:

$$\psi_{i,Aug}(\beta) = \frac{\delta_i}{\hat{\pi}(v_i)} S(v_i, \beta) + \left(1 - \frac{\delta_i}{\hat{\pi}(v_i)}\right) \hat{m}(z_i, \gamma), \quad i = 1, \dots, n, \quad (16)$$

where $m(z_i, \gamma) = E(S(v_i, \beta) | Z_i = z_i, \delta = 0)$. In a similar way to Kim and Yu (2011) and the same arguments used in Niu et al. (2014), we can estimate $m(z_i, \gamma)$ semiparametrically as follow:

$$\hat{m}(z_i, \gamma) = \frac{\sum_{j=1}^n \delta_j S(z_j, u_j, \beta) e^{\gamma c(u_j)} K_h(z_j - z_i)}{\sum_{j=1}^n \delta_j e^{\gamma c(u_j)} K_h(z_j - z_i)}, \quad i = 1, \dots, n. \quad (17)$$

Also we can rewrite $\psi_{i,Aug}(\beta)$ based on the score function $S(v_i, \beta)$, as follow:

$$\psi_{i,Aug}(\beta) = \sum_{j=1}^n W_{ij,0} S(v_i, \beta), \quad (18)$$

where by considering δ_{ij} as a Kronecker multiplier,

$$W_{ij,0} = \delta_j \left\{ \frac{\delta_{ij}}{\hat{\pi}(v_j)} + \left(1 - \frac{\delta_i}{\hat{\pi}(v_i)}\right) \frac{e^{\gamma c(u_j)} K_h(z_j - z_i)}{\sum_{j=1}^n \delta_j e^{\gamma c(u_j)} K_h(z_j - z_i)} \right\}. \quad (19)$$

Therefore, the augmented empirical log-likelihood ratio (AELR) can be defined as follow:

$$l_{Aug}(\beta) = -2sup \left\{ \sum_{i=1}^n \log(np_i) | p_i > 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \psi_{i,Aug}(\beta) = 0 \right\}. \quad (20)$$

By Lagrange method it is easy to obtain that the AELR function is as:

$$l_{Aug}(\beta) = 2 \sum_{i=1}^n \log(1 + \lambda \psi_{i,Aug}(\beta)). \quad (21)$$

Moreover, we can rewrite the AELR function based on the score function as follow:

$$l_{Aug}(\beta) = 2 \sum_{i=1}^n \log \left(1 + \sum_{j=1}^n W_{ij,0} \lambda^T S(z_j, u_j, \beta) \right). \quad (22)$$

The empirical maximum likelihood estimator of β based on the augmented method can be obtained by maximizing $-l_{Aug}(\beta)$ over all domain of β and λ . Based on the augmented method, we estimate β jointly with Lagrange multiplier by using the following two estimating equations:

$$\begin{cases} T_{1,Aug} = \frac{1}{n} \sum_{i=1}^n \frac{\sum_{j=1}^n W_{ij,0} S(z_j, u_j, \beta)}{1 + \sum_{j=1}^n W_{ij,0} \lambda^T S(z_j, u_j, \beta)} = 0 \\ T_{2,Aug} = \frac{1}{n} \sum_{i=1}^n \frac{\sum_{j=1}^n W_{ij,0} \lambda^T h(x_j) h^T(x_j)}{1 + \sum_{j=1}^n W_{ij,0} \lambda^T S(z_j, u_j, \beta)} = 0. \end{cases} \quad (23)$$

2.3 Empirical likelihood inference based on the improved augmented method

Again we consider augmented method, but we estimate mean score function, $m(Z, \gamma)$ as follow:

$$\hat{m}(z_i, \gamma) = \frac{\sum_{j=1}^n \delta_j S(z_i, u_j, \beta) e^{\gamma c(u_j)} K_h(z_j - z_i)}{\sum_{j=1}^n \delta_j e^{\gamma c(u_j)} K_h(z_j - z_i)}, \quad i = 1, \dots, n. \quad (24)$$

The difference between estimators given in Eqs. (17) and (24) is obvious where in the above estimator the observed variables of estimators remain without change (see z_i in $S(\cdot)$ which is the observed part of the i th individual and u_i is the missing part of the i th individual). This method is used by Creemers et al. (2011) to estimate parameters of a linear model with missing at random covariates. Also, in their studies, this method had a good performance. We call this method, Improved Augmented (IA) method. It is easy to obtain that the improved augmented empirical log-likelihood ratio (IAELR) is as follow:

$$l_{IA}(\beta) = 2 \sum_{i=1}^n \log \left(1 + \sum_{j=1}^n W_{ij,0} \lambda^T S(z_i, u_j, \beta) \right). \quad (25)$$

The empirical maximum likelihood estimator of β based on the IA method can be obtained by maximizing $-l_{IA}(\beta)$ over all domain of β and λ . Based on the IA method, we estimate β jointly with Lagrange multiplier by using the following two estimating equations:

$$\begin{cases} T_{1,IA} = \frac{1}{n} \sum_{i=1}^n \frac{\sum_{j=1}^n W_{ij,0} S(z_i, u_j, \beta)}{1 + \sum_{j=1}^n W_{ij,0} \lambda^T S(z_i, u_j, \beta)} = 0 \\ T_{2,IA} = \frac{1}{n} \sum_{i=1}^n \frac{\sum_{j=1}^n W_{ij,0} \lambda^T h(x_{i,obs}, x_{j,mis}) h^T(x_{i,obs}, x_{j,mis})}{1 + \sum_{j=1}^n W_{ij,0} \lambda^T S(z_j, u_j, \beta)} = 0, \end{cases} \quad (26)$$

where X_{obs} is the fully observed vectors of covariates and X_{mis} is the partly missed vector of covariates.

3 Theoretical results

In this section, we will study asymptotic properties of the three discussed methods. We will determine two theorems in the cases that tilting parameter is known and unknown. We will see asymptotic normality of proposed estimators and asymptotic convergence of the empirical log-likelihood ratios to a chi-square distribution. Also, we have assumed that the regularity conditions given in the Appendix hold.

3.1 Asymptotic inference with known tilting parameter

In this section, we assume that γ is known. For simplicity, we define the following notations:

$$A = E(\pi^{-1}(Z, U) h(X) h^T(X) \epsilon^2 + (1 - \pi^{-1}(Z, U)) E^2(h(X) \epsilon | Z = z, \delta = 0)), \\ T = E(h(X) h^T(X)), \quad B = E(\pi^{-1}(Z, U) h(X) h^T(X) \epsilon^2).$$

Theorem 1 Suppose the regularity conditions given in the Appendix hold. For true values of β and known tilting parameter, we have:

$$\begin{aligned} \text{(a)} \quad & \sqrt{n}(\hat{\beta}_{IPW} - \beta) \xrightarrow{D} N(0, T^{-1} A T^{-1}), \\ & \sqrt{n}(\hat{\beta}_{Aug} - \beta) \xrightarrow{D} N(0, T^{-1} A T^{-1}), \\ & \sqrt{n}(\hat{\beta}_{IA} - \beta) \xrightarrow{D} N(0, T^{-1} A T^{-1}), \\ \text{(b)} \quad & \hat{l}_{IPW}(\beta, \gamma) \xrightarrow{D} \rho_{1,IPW} \chi_{1,1}^2 + \rho_{2,IPW} \chi_{1,2}^2 + \cdots + \rho_{d,IPW} \chi_{1,d}^2 \\ & \hat{l}_{Aug}(\beta, \gamma) \xrightarrow{D} \chi_d^2, \end{aligned}$$

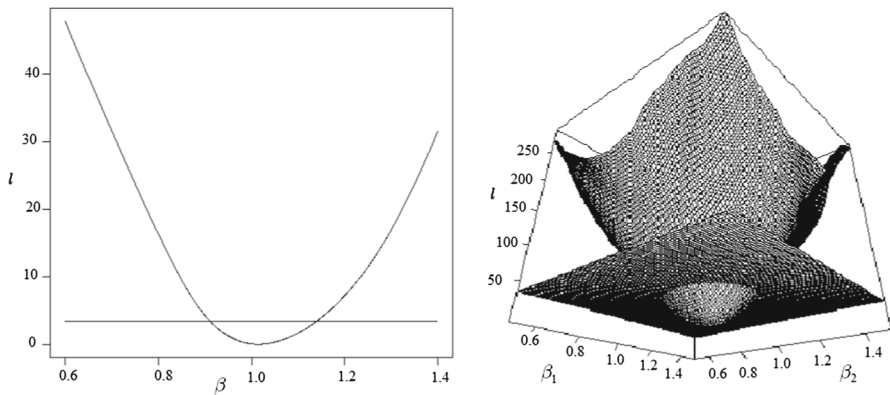


Fig. 1 Confidence regions in one parameter (left plot) and two parameters (right plot) cases

$$\hat{l}_{IA}(\beta, \gamma) \xrightarrow{D} \chi_d^2,$$

where $\chi_{1,i}^2$'s, $i = 1, \dots, d$ are the independent variables from the chi-square distribution with one degree of freedom, $\rho_{i,IPW}$'s, $i = 1, \dots, d$ are the eigenvalues of $B^{-1}A$ and χ_d^2 is the random variable from the chi-square distribution with d degrees of freedom.

We can use Theorem 1(a) to construct a confidence region for parameters of general linear model based on the different pivots and we can test the hypothesis such as $H_0 : \beta = \beta_0$ at level α . We call the inferences based on Theorem 1(a), normal approximation (NA) inferences. The part (b) of Theorem 1 gives the empirical log-likelihood ratio properties of the mentioned methods. We can use Theorem 1(b) to inference about parameters of the general linear model without needing to have any pivots where an approximate $100(1 - \alpha)\%$ confidence region for β consist of all possible β_0 's for which the null hypothesis $H_0 : \beta = \beta_0$ would not be rejected at level α . The approximate $100(1 - \alpha)\%$ confidence region by IPW method can be constructed by $CI_{IPW} = \{\beta | \hat{l}_{IPW}(\beta, \gamma) \leq \sum_{i=1}^d \rho_{i,IPW} \chi_{1,i}^2(1 - \alpha)\}$, where $\chi_{1,i}^2(1 - \alpha)$'s, $i = 1, \dots, d$ are the $(1 - \alpha)$ quantile of the Chi-square distribution with one degree of freedom. Also, the confidence region for parameters of the other methods can be constructed in the same way. To have a better perception about the confidence regions without using any pivots by the empirical log-likelihood ratio method, see also Fig. 1. The left curve in Fig. 1 shows the empirical log-likelihood ratio for the general linear model with one parameter that cut off by the vertical line, $l = \sum_{i=1}^d c_i \chi_{d,i}^2(1 - \alpha)$. The cut points give the beginning and the end of the approximate $100(1 - \alpha)\%$ confidence interval. Also, consider the right shape in Fig. 1 where shows the empirical log-likelihood ratio for the general linear model with two parameters. The confidence region obtained by mapping of the intersection between the empirical log-likelihood shape and the plane, $l = \sum_{i=1}^d c_i \chi_{d,i}^2(1 - \alpha)$ to coordinate plane of β_1 and β_2 . Where c_i 's and d are same as to those defined in Theorem 1.

The proof of Theorem 1 is given in the Appendix. Also, appropriate and consistent estimators of unknown statistics are given in the Appendix.

3.2 Asymptotic inference with unknown tilting parameter

Naturally, in statistical inferences, tilting parameter is unknown and we have to estimate it. Using validation sample is a common way to estimate tilting parameter. We estimate tilting parameter from the following consistent estimating equations,

$$\sum_{i=1}^n r_i (1 - \delta_i) (c(u_i) - \hat{m}^*(z_i, \gamma)) = 0, \quad (27)$$

where $m^*(z_i, \gamma) = E(c(U)|z_i, \delta = 0)$, r_i is an indicator function which takes value 1 if the unit i , $i = 1, \dots, n$ belongs to the follow-up sample and otherwise it is equal to 0. The validation sample is randomly selected from set of nonrespondents and the response are obtained for all the elements in validation sample. Also, the validation sample size is a pre-specified proportion of dataset, n .

One can estimate tilting parameter by using the conditions such as $E(\frac{\delta}{\pi(X, Y)}|x, y) = 1$. However, this condition gives an unbiased estimation of tilting parameter but estimation of tilting parameter with this condition is not appropriate because its variance is big and cause to bad results. Also, the bad performance of considering this condition to estimate the tilting parameter was obvious in our pre-simulation studies. For this reason, it seems using the mentioned method in Eq. (27) is necessary to estimate the tilting parameter.

For simplicity, we define the following notations,

$$\begin{aligned} M &= E(r(1 - \delta)(E(c^2(U)|Z = z, \delta = 0) - m^{*2}(Z, \gamma))), \\ H &= E((1 - \pi(U, Z))(c(U) - m^*(Z, \gamma))(S(U, Z, \beta) - m(Z, \gamma))), \\ \eta &= \frac{\delta}{\pi(V)} S(Z, U, \beta) + \left(1 - \frac{\delta}{\pi(V)}\right) E(S(Z, U, \beta)|z, \delta = 0) \\ &\quad + HM^{-1}[(1 - \delta)r - \delta E(r|\delta = 0)(\pi^{-1}(V) - 1)](c(U) - m^*(Z, \beta)), \end{aligned}$$

and define $\tilde{A}$ as the variance of random variable η . Now, we can construct the following theorem when tilting parameter is unknown.

Theorem 2 Suppose the conditions given in the Appendix hold. For true values of β and $\hat{\gamma}$ from Eq. (27), we have:

$$\begin{aligned} \text{(a)} \quad & \sqrt{n}(\hat{\beta}_{IPW} - \beta) \xrightarrow{D} N(0, T^{-1} \tilde{A} T^{-1}) \\ & \sqrt{n}(\hat{\beta}_{Aug} - \beta) \xrightarrow{D} N(0, T^{-1} \tilde{A} T^{-1}), \\ & \sqrt{n}(\hat{\beta}_{IA} - \beta) \xrightarrow{D} N(0, T^{-1} \tilde{A} T^{-1}). \\ \text{(b)} \quad & \hat{l}_{IPW}(\beta, \gamma) \xrightarrow{D} \rho_{1,IPW} \mathcal{X}_{1,1}^2 + \rho_{2,IPW} \mathcal{X}_{1,2}^2 + \dots + \rho_{d,IPW} \mathcal{X}_{1,d}^2, \end{aligned}$$

$$\begin{aligned}\hat{l}_{Aug}(\beta, \gamma) &\xrightarrow{D} \rho_{1,Aug} \chi_{1,1}^2 + \rho_{2,Aug} \chi_{1,2}^2 + \cdots + \rho_{d,Aug} \chi_{1,d}^2, \\ \hat{l}_{IA}(\beta\gamma) &\xrightarrow{D} \rho_{1,IA} \chi_{1,1}^2 + \rho_{2,IA} \chi_{1,2}^2 + \cdots + \rho_{d,IA} \chi_{1,d}^2.\end{aligned}$$

The weights $\rho_{i,IPW}$, $\rho_{i,Aug}$ and $\rho_{i,IA}$, $i = 1, \dots, d$ are the eigenvalues of $B^{-1} \tilde{A}$, $A^{-1} \tilde{A}$ and $A^{-1} \tilde{A}$, respectively. In this theorem, the empirical log-likelihood distribution of IPW method is similar to results of the first theorem. In this case, the coefficients of chi-square distribution are different. Moreover, the empirical log-likelihood distributions of augmented methods are different in comparison to the results of Theorem 1. This difference follows by the second lemma which is specified in the Appendix. The proof of Theorem 2 and appropriate consistent estimators of unknown functions is given with details in the Appendix.

4 Simulation study

In this section, we will study the performances of the three mentioned methods. We divide the simulation study to two sections, wherein Sect. 4.1, we have studied a general linear model with one covariate and in Sect. 4.2, we have studied a general linear model with two covariates. To see performances of the proposed methods, we have implemented several cases of missing data mechanism by different rates of the missingness in variables. Also, we have considered several sample sizes to see the sample size effect on inferences. Moreover, the confidence level to construct confidence regions is 95% for all methods.

4.1 General linear model with one covariates

For the general linear model of Eq. (1), suppose $h(X) = 1 + X$, where, X is a random variable from normal distribution $N(0, 1)$, ϵ is a random variable from normal distribution $N(0, 0.25)$ and parameter of the general linear model is equal to 1.

Study 1. Missingness in response

Assume that covariate X is fully observed and missing data occurred in the response variable Y . The following three missing patterns is given to generate missing data in the response variable:

- Case 1. $\pi_1(x_i, y_i) = 1/(1 + e^{-(0.5x_i + 0.5y_i + 0.8)})$,
- Case 2. $\pi_2(x_i, y_i) = 1/(1 + 0.5(0.6 + |x_i|)e^{-x_i y_i})$,
- Case 3. $\pi_3(x_i, y_i) = 1/(1 + e^{-(0.6x_i + 0.6y_i + 1.6)})$,

by which we will have approximately 25%, 25% and 15% rates of missing data in Y , respectively. Case 1 and Case 3 both are logistic models but with different percentage of missing data which help us to see the effect of missing rates effects on inferences. Also, the second pattern is a nonlogistic model with the same rate of missingness to the Case 1 and used to see performances of proposed methods under the nonlogistic patterns.

The Gaussian function is used as a kernel function and bandwidth parameter h is estimated by Cross Validation method. 30% rate of the follow-up sample is used to estimate γ by Eq. (27) where $c(u) = u$ is chosen.

Table 1 constructed by 3000 Monte Carlo repetitions. Where, absolute biases (AB) of estimators, mean square errors (MSE) of estimators, internal range of quantile (IQR) for 3000 Monte Carlo repetition, coverage probability by normal (CPN) approximation, coverage probability by empirical log-likelihood (CPEL) approximation, average widths of confidence intervals by normal (AWCIN) approximation and average widths of confidence intervals by empirical log-likelihood (AWCIEL) approximation, for three methods IPW, Aug and IA, are given in Table 1.

From Table 1, by increasing the sample size all three methods have progressive performances. Also, improvement in results can be seen by decreasing the rate of missingness. Method IA does not show a good performance for the high rate of missingness, but its performance improves when the sample size increases. Two other methods also have approximately the same performances. Where IPW method have an appropriate coverage probability and it seems that the length of the confidence intervals for the Aug method are lower than those of the IPW method. Overall in this study, all three methods lead to good performances unless IA method when we have the high rate of missingness and small sample size.

Study 2. Missingness on covariate

Assume that response variable Y is fully observed and some individuals of covariate X are missing by the above patterns. Also, in this case, we apply the similar argument used in Study 1. Therefore, the rate of missingness will be the same as those of the last study. Results of this study are given in Table 2.

From Table 2, we can conclude that results are the same as Study 1. But in this case, IA method have coverage probability at least as good as IPW method. When missing values happen in covariate variable, Aug method has a better performance in comparison with two other methods. Also, in this case, IA method has a better performance in comparison with that of Study 1. Overall in this study, all three methods have good performances.

4.2 General linear model with two covariates

Study 3. Missingness on the response or a covariate

For general linear model of Eq. (1), suppose $h(x) = (x_1^2, \log(4 + x_2))$, where X_1 and X_2 are random variables from Uniform distributions $U(0, 1)$ and $U(-1, 1)$, respectively. Moreover, it is assumed that $\beta = (2, 1)$.

In this study we assume that missingness in response or covariates follows from one of the following MNAR mechanisms,

$$\text{Case 4. } \pi_4(x_{1,i}, x_{2,i}, y_i) = 1/(1 + e^{-(0.6x_{1,i} - x_{2,i} + 0.6y_i)}),$$

$$\text{Case 5. } \pi_5(x_{1,i}, x_{2,i}, y_i) = 1/(1 + |x_{1,i} + x_{2,i}|e^{-0.75x_{1,i}y_i}).$$

In both cases, 20.4% of data will be missing approximately. The independent Gaussian kernels are used to estimate unknown functions where the rate of follow-up is 30% where $c(u) = u$ is chosen.

Table 1 Effect of sample sizes, missingness rates, and missingness mechanisms on inferences about parameters of a general linear model with one covariate when the response variable is missing not at random

π	n	Method	AB	MSE	IQR	CPN	CPEL	AWCIN	AWCIEL
π_1	50	IPW	0.0449	0.0032	0.0757	0.9463	0.9447	0.24219	0.2482
		Aug	0.0458	0.0034	0.0770	0.9423	0.9390	0.2343	0.2360
		IA	0.0523	0.0046	0.0809	0.9217	0.9220	0.2345	0.2447
	100	IPW	0.0315	0.0016	0.0536	0.9533	0.9547	0.1608	0.1636
		Aug	0.0319	0.0016	0.0540	0.9517	0.9527	0.1582	0.1619
		IA	0.0353	0.0020	0.0569	0.9327	0.9353	0.1583	0.1654
	150	IPW	0.0261	0.0011	0.0445	0.9503	0.9473	0.1291	0.1304
		Aug	0.0265	0.0011	0.0447	0.9430	0.9443	0.1275	0.1306
		IA	0.0290	0.0013	0.0473	0.9223	0.9233	0.1276	0.1326
π_2	50	IPW	0.0437	0.0030	0.0717	0.9380	0.9377	0.2091	0.2130
		Aug	0.0438	0.0030	0.0724	0.9410	0.9390	0.2136	0.2163
		IA	0.0444	0.0031	0.0731	0.9397	0.9357	0.2135	0.2170
	100	IPW	0.0309	0.0015	0.0499	0.9380	0.9370	0.1493	0.1509
		Aug	0.0309	0.0015	0.0496	0.9413	0.9410	0.1522	0.1525
		IA	0.0314	0.0016	0.0504	0.9400	0.9370	0.1522	0.1529
	150	IPW	0.0260	0.0011	0.0417	0.9393	0.9360	0.1222	0.1229
		Aug	0.0258	0.0011	0.0418	0.9440	0.9367	0.1244	0.1238
		IA	0.0111	0.0263	0.0401	0.9397	0.9307	0.1244	0.1240
π_3	50	IPW	0.0419	0.0028	0.0709	0.9523	0.9537	0.2177	0.2223
		Aug	0.0424	0.0028	0.0716	0.9453	0.9467	0.2136	0.2149
		IA	0.0452	0.0033	0.0744	0.9400	0.9390	0.2138	0.2117
	100	IPW	0.0298	0.0014	0.0523	0.9533	0.9520	0.1517	0.1540
		Aug	0.0300	0.0014	0.0525	0.9483	0.9460	0.1496	0.1521
		IA	0.0326	0.0016	0.0547	0.9347	0.9370	0.1497	0.1550
	150	IPW	0.0246	0.0009	0.0402	0.9463	0.9457	0.1211	0.1219
		Aug	0.0247	0.0010	0.0408	0.9447	0.9433	0.1198	0.1218
		IA	0.0261	0.0011	0.0401	0.9317	0.9363	0.1198	0.1232

Where, 3000 Monte Carlo repetitions are used to estimate some criteria (*AB* absolute bias, *MSE* mean square error, *IQR* internal range of quantile, *CPN* coverage probability by normal approximation, *CPEL* coverage probability by empirical log-likelihood approximation, *AWCIN* average width of confidence interval by normal approximation, *AWCIEL* average width of confidence interval with empirical log-likelihood approximation)

Table 3 is constructed by 3000 Monte Carlos repetitions. In the first part of Table 3, we have assumed that missingness happened in the response variable and covariates are fully observed and in the other part of Table 3, we have assumed that missingness happened in the covariate X_1 and the other two variables are fully observed. The coverage probabilities of intervals obtained by different methods are shown by *CPN* and *CPEL* for normal approximation and empirical log-likelihood approximation, respectively. Also, we have used the *MSE* to see the performance of estimators of different methods.

Table 2 Effect of sample sizes, missingness rates, and missingness mechanisms on inferences about parameters of a general linear model with one covariate when covariate variable is missing not at random

π	n	Method	AB	MSE	IQR	CPN	CPEL	AWCIN	AWCIEL
π_1	50	IPW	0.0436	0.0030	0.0742	0.9500	0.9487	0.2444	0.2507
		Aug	0.0437	0.0030	0.0740	0.9447	0.9490	0.2390	0.2553
		IA	0.0440	0.0031	0.0736	0.9397	0.9467	0.2394	0.2584
	100	IPW	0.0306	0.0015	0.0526	0.9567	0.9570	0.1613	0.1645
		Aug	0.0307	0.0015	0.0525	0.9513	0.9607	0.15868	0.1690
		IA	0.0308	0.0015	0.0519	0.9513	0.9537	0.1587	0.1708
	150	IPW	0.0253	0.0010	0.0430	0.9570	0.9563	0.1258	0.1339
		Aug	0.0256	0.0010	0.0436	0.9507	0.9593	0.1258	0.1328
		IA	0.0258	0.0010	0.0435	0.9536	0.9593	0.1258	0.1339
π_2	50	IPW	0.0424	0.0028	0.0708	0.9447	0.9453	0.2184	0.2230
		Aug	0.0425	0.0029	0.0710	0.9460	0.9527	0.2198	0.2269
		IA	0.0428	0.0029	0.0703	0.9440	0.9493	0.2194	0.2184
	100	IPW	0.0299	0.0014	0.0506	0.9473	0.9490	0.1509	0.1502
		Aug	0.0301	0.0014	0.0505	0.9483	0.9527	0.1522	0.1590
		IA	0.0303	0.0014	0.0510	0.9487	0.9533	0.1521	0.1565
	150	IPW	0.0245	0.0009	0.0416	0.9553	0.9533	0.1223	0.1229
		Aug	0.0246	0.0009	0.0418	0.9550	0.9563	0.1234	0.1257
		IA	0.0247	0.0009	0.0419	0.9527	0.9523	0.1233	0.1234
π_3	50	IPW	0.0414	0.0027	0.0700	0.9510	0.9507	0.2223	0.2271
		Aug	0.0417	0.0027	0.0700	0.9467	0.9490	0.2179	0.2298
		IA	0.0418	0.0027	0.0703	0.9476	0.9523	0.2180	0.2216
	100	IPW	0.0296	0.0013	0.0514	0.9563	0.9547	0.1522	0.1545
		Aug	0.0297	0.0013	0.0517	0.9523	0.9570	0.1496	0.1569
		IA	0.0296	0.0013	0.0518	0.9553	0.9580	0.1497	0.1582
	150	IPW	0.0241	0.0009	0.0398	0.9467	0.9417	0.1214	0.1223
		Aug	0.0243	0.0009	0.0408	0.9407	0.9457	0.1196	0.1245
		IA	0.0243	0.0009	0.0412	0.9400	0.9463	0.1196	0.1255

Where 3000 Monte Carlo repetitions are used to estimate some criteria (*AB* absolute bias, *MSE* mean square error, *IQR* internal range of quantile, *CPN* coverage probability by normal approximation, *CPEL* coverage probability by empirical log-likelihood approximation, *AWCIN* average width of confidence interval by normal approximation, *AWCIEL* average width of confidence interval with empirical log-likelihood approximation)

$$MSE = \frac{1}{t} \sum_{i=1}^t \left(\left(\frac{\hat{\beta}_{1,i} - \beta_1}{\beta_1} \right)^2 + \left(\frac{\hat{\beta}_{2,i} - \beta_2}{\beta_2} \right)^2 \right),$$

where t is the number of repetitions of the Monte Carlo algorithm. In addition, $\hat{\beta}_{1,i}$ and $\hat{\beta}_{2,i}$ are the estimates of general linear model parameters in the i th repetition of the Monte Carlo algorithm.

Table 3 Effect of sample sizes, missingness rates, and missingness mechanisms on inferences about parameters of a general linear model with two covariates when Y or X_1 contain missing data

Missed variable	n	Method	100			150		
			MSE	CPN	CPEL	MSE	CPN	CPEL
Y	π_4	IPW	0.0131	0.9453	.9553	0.0089	0.9523	0.9590
		Aug	0.0134	0.9210	0.9623	0.0092	0.9230	0.9687
		IA	0.0136	0.9153	0.9587	0.0093	0.9220	0.9700
	π_5	IPW	0.0127	0.9513	0.9607	0.0085	0.9552	0.9580
		Aug	0.0135	0.9247	0.9660	0.0091	0.9237	0.9657
		IA	0.0134	0.9253	0.9650	0.0091	0.9230	0.9667
X_1	π_4	IPW	0.0129	0.9417	.9583	0.0088	0.9503	0.9613
		Aug	0.0120	0.9420	0.9503	0.0079	0.9550	0.9587
		IA	0.0114	0.9477	0.9560	0.0074	0.9583	0.9660
	π_5	IPW	0.0119	0.9553	0.9667	0.0079	0.9560	0.9653
		Aug	0.0115	0.9537	0.9540	0.0076	0.9537	0.9527
		IA	0.0111	0.9553	0.9590	0.0074	0.9557	0.9563

Where 3000 Monte Carlo repetitions are used to estimate some criteria (*MSE* rational mean square error, *CPN* coverage probability by normal approximation, *CPEL* coverage probability by empirical log-likelihood approximation)

From Table 3, when the response variable contains missing data, IPW method has slightly better performance in comparison to two other methods. However, in this case, the CPN of the IPW method is more realistic than the CPN of the augmented methods. The other criteria in terms of performance are not too different. On the other hand, when covariate X_1 contains missing data, the augmented methods have better performance in comparison to that of IPW method in terms of Bias and MSE. In missing at covariate case, one can prefer the performance of the IA method in comparison to those of the IPW and Aug methods. Moreover, by increasing the sample size performance of all methods improve.

The augmented method is used by Wang and Wang (2001) and IA method is introduced by Creemers et al. (2011) to inference about the linear model with missing on covariates by MAR mechanism. Good performance of these methods is obvious by results of Table 2 and especially the second part of Table 3 when missing data happens in covariates. Moreover, when missing happens in covariates, by increasing the number of covariates their performance improved. However, when missing data happens in the response variable, the performance of IA method is not as good as the performance of Aug and IPW methods.

Study 4. Missingness on the response and a covariate

For general linear model of Eq. (1), suppose $h(x) = (x_1, x_2^2 + \log(1 + x_2))$, where X_1 is a random variable from Normal distribution $N(0, 1)$ and X_2 is a random variable from Uniform distribution $U(-1, 1)$. Moreover, it is assumed that $\beta = (1, 1)$.

In this study, we assume that missing data in response and covariate x_1 are generated by one of the following mechanisms:

Case 6. $\pi_6(x_{1,i}, x_{2,i}, y_i) = 1/(1 + e^{-(0.5x_{1,i} + 0.5x_{2,i} + 0.25y_i)})$,

Case 7. $\pi_7(x_{1,i}, x_{2,i}, y_i) = 1/(1 + 0.5|x_{2,i}|e^{0.25-x_{1,i}y_i})$,

Case 8. $\pi_8(x_{1,i}, x_{2,i}, y_i) = 1/(1 + e^{-(1.75 + 0.25x_{2,i})})$.

where π_6 and π_8 are both logistic functions. π_6 and π_7 follow both from MNAR mechanisms and π_8 follows from MAR mechanism. By the above missing mechanisms, we will have 15% missing data on response variable and covariate X_1 . Moreover, by the above mechanisms, we have assumed that for i th member of the sample, Y and X_1 are both missed or both observed. In addition, we have considered the different functions of $c(u)$ as follow:

- $c_1(u) = x_1 + y$,
- $c_2(u) = x_1$,
- $c_3(u) = y$.

These different functions are considered to see the performance of different logistic function assumption for missingness probability because $c(\cdot)$ function plays an important role in the estimation of unknown function and specially in the estimation of tilting parameter. Moreover, the rate of follow-up sample is still 30%. Results of this study are given in Table 4.

From Table 4, one can conclude that the coverage probability of the IA method is closed to 0.95 when the missing mechanism has the logistic form. However, by going away from logistic function, its coverage probability is less than of our expectation i.e. 0.95. Moreover, IPW and Aug method show good performances in many cases but the coverage probability of IPW method based on π_6 and π_8 is more than 0.95. This is more obvious when the missing mechanism is the sixth mechanism. On the other hand, the coverage probability of the Aug method decreased similarly to the coverage probability of the IA method under the seventh mechanism.

By comparing three rows of Table 4, we can say that applying different $c(u)$ does not more effects in conclusions. However, we can say if $c(u)$ chosen to be function of the covariates, it can improve the performance of the augmented methods.

5 Real data study

In this section, the evaporative resistance of male Chinese ethnic clothing dataset is used to fit a general linear model. This dataset is given in Wang et al. (2016) where the air layer distribution and evaporative resistances of 39 sets of male Chinese ethnic clothing were investigated using a sweating thermal manikin and the three-dimensional body scanning technique. The methods of measuring the data with more details are given in Wang et al. (2016).

For this dataset, we investigate the relationship between the clothing total evaporative resistance (response variable Y) and the total volume of the air entrapped inside the clothing (covariate variable X). Wang et al. (2016) have fitted some general linear model for these variables. They have considered polynomial models with three parameters but we consider some different model with two parameters. We have standardized the data to have simple calculations. Moreover, this dataset has been fully

Table 4 Effect of sample sizes, missingness rates, and missingness mechanisms on inferences about parameters of a general linear model with two covariates when Y and X_1 contain missing data

$c(u)$	n		100			150		
	π	Method	MSE	CPN	CPEL	MSE	CPN	CPEL
$c_1(u) = x_1 + y$	π_6	IPW	0.0077	0.9667	0.9793	0.0050	0.9750	0.9865
		Aug	0.0079	0.9663	0.9553	0.0050	0.9720	0.9627
		IA	0.0172	0.9323	0.9350	0.0120	0.9320	0.9480
	π_7	IPW	0.0077	0.9500	0.9687	0.0053	0.9483	0.9683
		Aug	0.0078	0.9477	0.9310	0.0053	0.9487	0.9213
		IA	0.0124	0.9283	0.9107	0.0121	0.9317	0.9213
	π_8	IPW	0.0068	0.9547	0.9673	0.0043	0.9643	0.9733
		Aug	0.0067	0.9540	0.9490	0.0043	0.9650	0.9547
		IA	0.0098	0.9336	0.9490	0.0062	0.9417	0.9527
$c_2(u) = x_1$	π_6	IPW	0.0076	0.9640	0.9780	0.0049	0.9730	0.9867
		Aug	0.0078	0.9627	0.9517	0.0050	0.9703	0.9630
		IA	0.0173	0.9470	0.9387	0.0128	0.9573	0.9533
	π_7	IPW	0.0078	0.9522	0.9707	0.0052	0.9540	0.9707
		Aug	0.0078	0.9510	0.9336	0.0053	0.9550	0.9287
		IA	0.0170	0.9297	0.9213	0.0124	0.9393	0.9273
	π_8	IPW	0.0067	0.9510	0.9653	0.0044	0.9653	0.9743
		Aug	0.0067	0.9520	0.9463	0.0044	0.9643	0.9547
		IA	0.0097	0.9363	0.9510	0.0062	0.9460	0.9557
$c_3(u) = y$	π_6	IPW	0.0077	0.9670	0.9776	0.0051	0.9717	0.9823
		Aug	0.0076	0.9670	0.9497	0.0051	0.9710	0.9557
		IA	0.0159	0.9183	0.9110	0.0119	0.9117	0.9203
	π_7	IPW	0.0078	0.9493	0.9660	0.0053	0.9483	0.9683
		Aug	0.0078	0.9470	0.9273	0.0053	0.9487	0.9213
		IA	0.0168	0.9230	0.9117	0.0112	0.9317	0.9213
	π_8	IPW	0.0077	0.9567	0.9643	0.0044	0.9620	0.9727
		Aug	0.0076	0.9550	0.9470	0.0044	0.9623	0.9520
		IA	0.0099	0.9357	0.9467	0.0064	0.9430	0.9483

Where 3000 Monte Carlo repetitions are used to estimate some criteria (*MSE* rational mean square error, *CPN* coverage probability by normal approximation, *CPEL* coverage probability by empirical log-likelihood approximation)

observed. Therefore, in order to see the performance of used methods, we generate missing data by the MNAR mechanism in covariate by the following mechanisms:

Case 9. $\pi_9(x_i, y_i) = 1/(1 + e^{-(0.75+0.5x_i+1.5x_i^2+0.5y_i)})$,

Case 10. $\pi_{10}(x_i, y_i) = 1/(1 + 0.5e^{-x_i y_i})$.

Based on the ninth mechanism, we will have approximately 24.2 % rate of missing data in the covariate variable and based on the tenth mechanism, we will have approximately 24.5 % rate of missing data in the covariate variable. We have considered the following general linear models:

Table 5 Inference about general linear model based on the standardized dataset of the evaporative resistance of male Chinese ethnic clothing. P values are estimated based on the normal approximation and the empirical log-likelihood approximation

Model	Method	π_9		π_{10}		$\hat{\beta}_1$	P value
		$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$		
1	IPW	0.0224	0.5790	4.7×10^{-4} (1.2×10^{-4})	-0.0091	0.6258	9.9×10^{-5} (4.1×10^{-16})
	Aug	0.0141	0.5794	4.7×10^{-4} (4.8×10^{-14})	-0.0012	0.6147	1.1×10^{-4} (5.1×10^{-14})
	IA	-0.0003	0.5856	4.7×10^{-4} (3×10^{-11})	0.0013	0.6047	9.9×10^{-5} (4.2×10^{-7})
2	IPW	-0.1018	0.1239	0.0419(0.0142)	-0.1562	0.1415	0.0260(0.0137)
	Aug	-0.1181	0.1252	0.0331(0.0240)	-0.1465	0.1389	0.0286(0.0235)
	IA	-0.1337	0.1263	0.0419(0.0385)	-0.1405	0.1355	0.0261(0.0808)
3	IPW	0.0513	-0.1402	0.3685(0.6044)	0.0131	-0.1396	0.4085(0.6360)
	Aug	0.0374	-0.1516	0.3666(0.6410)	0.0189	-0.1444	0.4125(0.6727)
	IA	0.0234	-0.1540	0.3686(0.5935)	0.0207	-0.1415	0.4059(0.6068)
4	IPW	0.8339	-0.1612	0.0010(0.0004)	0.9103	-0.1774	0.0012(0.0005)
	Aug	0.8275	-0.1595	0.0008(0.0005)	0.8912	-0.1719	0.0011(0.0004)
	IA	0.8462	-0.1675	0.0010(0.0005)	0.8788	-0.1700	0.0012(0.0004)

Where p values of the second method are given inside of the parenthesis

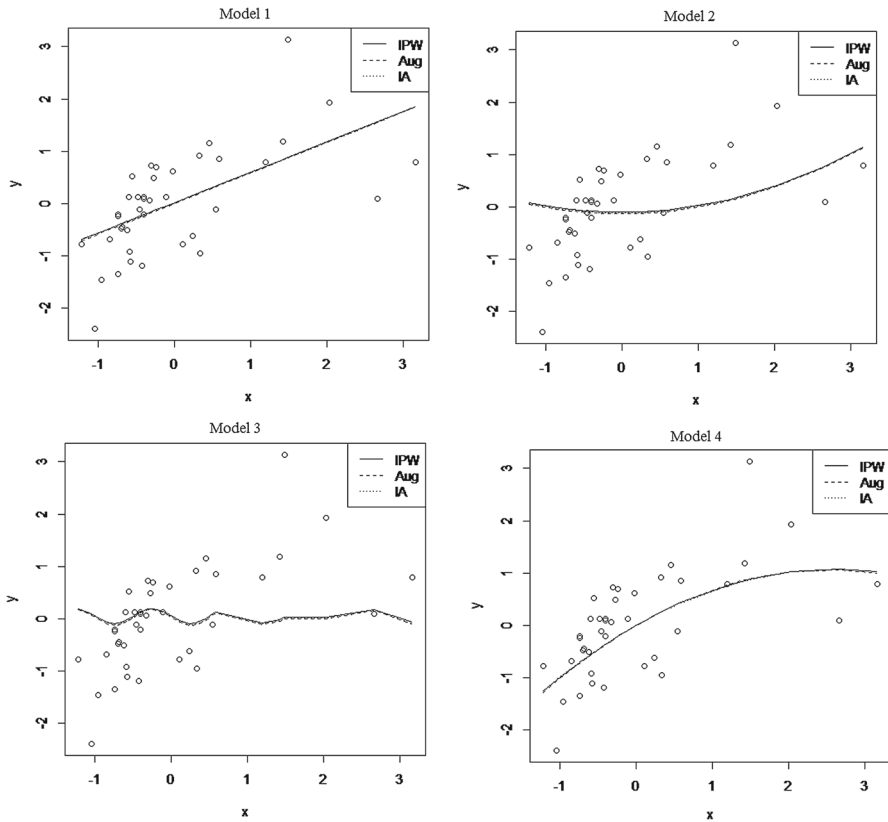


Fig. 2 Fitted models to the standardized dataset of the evaporative resistance of male Chinese ethnic clothing

Model 1: $y = \beta_0 + \beta_1 x$

Model 2: $y = \beta_0 + \beta_1 x^2$

Model 3: $y = \beta_0 + \beta_1 \sin(2\pi x)$

Model 4: $y = \beta_0 x + \beta_1 x^2$

In addition to estimation of parameters, we have tested $H_0 : \beta = 0$ assumption, where $\beta = (\beta_0, \beta_1)^T$. Results of this study are given in Table 5. The standardized data together with fitted models are given in Fig. 2 and the approximate 95% confidence regions for parameters of Model 4 are given in Fig. 3.

From Table 5, We can say that Models 2 and 3 cannot be appropriate to this dataset because we can not reject $H_0 : \beta = 0$ assumption at level 0.01. In addition, there is no reason to reject $H_0 : \beta = 0$ assumption at level 0.001 for Models 1 and 4 and they can be considered as the useful models. However, in this paper, our aim was not to find the best model but by our proposed model we have estimated parameters and we have tested some assumption about the parameters of the general linear model. The fitted models and the confidence regions of parameters based on model 4 are given in Figs. 2 and 3, respectively.

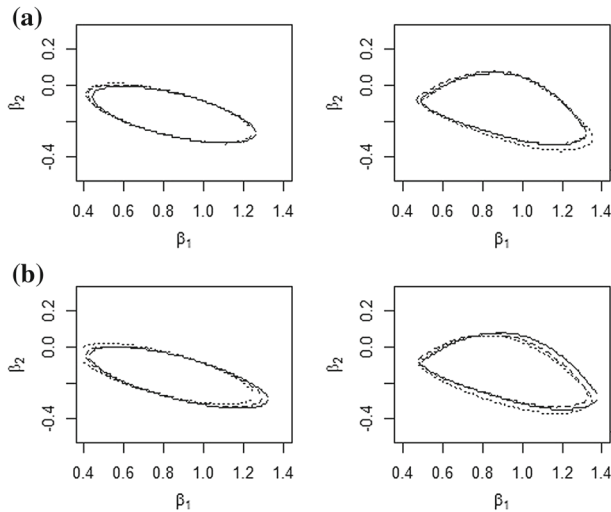


Fig. 3 Confidence regions for parameters of Model 4. The regions with a solid curve, a dotted curve, and a dashed curve show the IPW, IA, and Aug methods confidence regions, respectively. The plots in **a** show the confidence regions based on the ninth mechanism and the plots in **b** show the confidence regions based on the tenth mechanism. Moreover, the right shapes are based on the normal approximation and the left shapes are based on the empirical log-likelihood ratio method

From Fig. 3, the shapes of confidence regions are oval based on the normal approximation. When we use the empirical log-likelihood ratio to construct confidence regions, their shapes are similar to an ellipse.

6 Discussion

In our simulation studies, we have used the independent Gaussian kernel functions where unknown bandwidth parameters have been estimated by cross-validation method. The kernel function can be selected differently and it does not affect the inferences. But, choosing the bandwidth parameter is more critical than choosing the kernel function and it may affect the inferences. As mentioned, we have used cross-validation method where its good performances have been studied by many researchers.

The simulation study showed that Our three used methods have acceptable results. The IPW method has a more stable performance in comparison with the two other methods. On the other hand, IA method has fewer biases and MSE's for parameters in comparison with the other methods. However, our simulation study is based on the small sample size and by increasing sample size, we have expected an improvement in the performances of our methods.

Appendix

To achieve our conclusions, we assume that regularity conditions to be hold (Wang and Wang 2001). These conditions are as follow:

- (1) $\pi(v)$ is bounded and has partial derivatives up to order 2 almost surely.
- (2) Kernel function $k_h(\cdot)$, is continuous and is from order r . It is always at least from order 2.
- (3) The density function of Z , $f(z)$, exists and has bounded derivatives up to at least order 2.
- (4) The conditional expectations $E(S|Z = z)$ and $E(SS^T|Z = z)$ exist and have r continuous and bounded partially derivatives with respect to v .
- (5) For the score function S , $E(SS^T)$ exists and is positive definite.
- (6) $\eta_n = [nh^{2r} + (nh^{2d})^{-1}]$ converges to zero as n goes to infinity. Where d is the dimension of vector Z .
- (7) The matrices, A and B , defined in Sect. 3, are positive definite.

The following Lemma is necessary for proving mentioned theorems in Sect. 3. In any cases, we just proof parts (b) of lemmas and theorems with details. Proof of the other parts can be concluded in a similar manner to proof of part (b). However, the details of proof of part (a) are given anywhere which is necessary.

Lemma 1 *Under the regularity conditions, for known values of γ and true parameter β , we have*

$$(a) \quad \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{i,IPW}(\beta) \xrightarrow{D} N(0, A), \quad \frac{1}{n} \sum_{i=1}^n \psi_{i,IPW}(\beta) \psi_{i,IPW}^T(\beta) \xrightarrow{D} B, \quad (28)$$

$$(b) \quad \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{i,Aug}(\beta) \xrightarrow{D} N(0, A), \quad \frac{1}{n} \sum_{i=1}^n \psi_{i,Aug}(\beta) \psi_{i,Aug}^T(\beta) \xrightarrow{D} A, \quad (29)$$

$$(c) \quad \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{i,IA}(\beta) \xrightarrow{D} N(0, A), \quad \frac{1}{n} \sum_{i=1}^n \psi_{i,IA}(\beta) \psi_{i,IA}^T(\beta) \xrightarrow{D} A. \quad (30)$$

Proof For left hand side of part (b), we can write:

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{i,Aug}(\beta) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\delta_i}{\hat{\pi}_i(v_i)} S(v_i, \beta) + \left(1 - \frac{\delta_i}{\hat{\pi}_i(v_i)}\right) \hat{m}(z_i, \gamma) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\delta_i}{\hat{\pi}_i(v_i)} S(v_i, \beta) + \left(1 - \frac{\delta_i}{\hat{\pi}_i(v_i)}\right) m(z_i, \gamma) + o_P(1). \end{aligned}$$

The last equation follows from the fact, $\hat{m}(z_i, \gamma) = m(z_i, \gamma) + o_P(1)$ where its proof is given by many authors such as Kim and Yu (2011). Therefore, we can rewrite the above equation as follow:

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{i,Aug}(\beta) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\delta_i}{\pi_i(v_i)} S(v_i, \beta) + \left(1 - \frac{\delta_i}{\pi_i(v_i)}\right) m(z_i, \gamma) \end{aligned}$$

$$\begin{aligned}
& + \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{\delta_i}{\hat{\pi}_i(v_i)} - \frac{\delta_i}{\pi_i(v_i)} \right) (S(v_i, \beta) - m(z_i, \gamma)) \\
& + o_P(1) = J_1 + J_2 + o_P(1).
\end{aligned} \tag{31}$$

By using equal function of $\hat{\pi}(v_i)$ from Eq. (10), J_2 will be,

$$\begin{aligned}
J_2 &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \delta_i (1 - \hat{\alpha}(z_i) e^{\gamma c(u_i)} - 1 + \alpha(z_i) e^{\gamma c(u_i)}) (S(v_i, \beta) - m(z_i, \gamma)) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \delta_i e^{\gamma c(u_i)} (\hat{\alpha}(z_i) - \alpha(z_i)) (S(v_i, \beta) - m(z_i, \gamma)) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \delta_i e^{\gamma c(u_i)} (S(v_i, \beta) - m(z_i, \gamma)) \left(\frac{\sum_{j=1}^n (1 - \delta_j) K_h(z_i, z_j)}{\sum_{j=1}^n \delta_j e^{\gamma c(u_j)} K_h(z_i, z_j)} - \alpha(z_i) \right).
\end{aligned}$$

Also, the last equation follows from Eq. (9). It is easy to prove that:

$$\frac{1}{n} \sum_{j=1}^n \delta_j e^{\gamma c(u_j)} K_h(z_i, z_j) = \frac{1}{n} \sum_{j=1}^n \delta_j \left(\frac{1}{\pi(v_j)} - 1 \right) \alpha^{-1}(z_j) K_h(z_i, z_j).$$

Also, based on Eq. (5), by applying Strong Law of Large Numbers (SLLN) and some mathematical operations, we can conclude:

$$\frac{1}{n} \sum_{j=1}^n \delta_j \left(\frac{1}{\pi(v_j)} - 1 \right) \alpha^{-1}(z_j) K_h(z_i, z_j) = \alpha^{-1}(z_i) f(z_i) E(1 - \delta|z_i) + o_P(1),$$

where, $f(z_i)$ is the density function of Z_i . Therefore, we can write J_2 as follow,

$$\begin{aligned}
J_2 &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \delta_i e^{\gamma c(u_i)} (S(v_i, \beta) - m(z_i, \gamma)) \\
&\quad \times \left(\frac{1}{n} \frac{\sum_{j=1}^n ((1 - \delta_j) - \alpha(z_i) \delta_j e^{\gamma c(u_j)}) K_h(z_i, z_j)}{\alpha^{-1}(z_i) f(z_i) E(1 - \delta|z_i)} \right) + o_P(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \delta_i O(v_i) (S(v_i, \beta) - m(z_i, \gamma)) \\
&\quad \times \left(\frac{1}{n} \frac{\sum_{j=1}^n ((1 - \delta_j) - \alpha(z_i) \delta_j e^{\gamma c(u_j)}) K_h(z_i, z_j)}{f(z_i) E(1 - \delta|z_i)} \right) + o_P(1) \\
&= \frac{1}{n^{\frac{3}{2}}} \sum_{j=1}^n \sum_{i=1}^n \frac{\delta_i O(v_i) (S(v_i, \beta) - m(z_i, \gamma))}{f(z_i) E(1 - \delta|z_i)} \\
&\quad \times ((1 - \delta_j) - \alpha(z_i) \delta_j e^{\gamma c(u_j)}) K_h(z_i, z_j) + o_P(1).
\end{aligned}$$

For enough large sample sizes, by using SLLN for inner summation of the above equation, it is easy to prove that,

$$\begin{aligned} J_2 &= \frac{1}{\sqrt{n}} \sum_{j=1}^n ((1 - \delta_j) - \delta_j O(v_j)) E(S(V, \beta) - m(Z, \gamma) | z_j, \delta = 0) + o_P(1) \\ &= \frac{1}{\sqrt{n}} \sum_{j=1}^n ((1 - \delta_j) - \delta_j O(v_j)) (E(S(V, \beta) | z_j, \delta = 0) - m(z_j, \gamma)) \\ &\quad + o_P(1) = o_P(1). \end{aligned}$$

By combining J_1 and J_2 in Eq. (32), we will have:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{i,Aug}(\beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\delta_i}{\pi_i(v_i)} S(v_i, \beta) + \left(1 - \frac{\delta_i}{\pi_i(v_i)}\right) m(z_i, \gamma) + o_P(1).$$

It is very simple to get that $E(J_1 + J_2) = o_P(1)$ and $var(J_1 + J_2) = A$. Therefore, the left hand side of the part (b) of Lemma 1 can be concluded by the Central Limit Theorem (CLT). On the other hand, we can write the right hand side of the part (b) of Lemma 1 as:

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^n \psi_{i,Aug}(\beta) \psi_{i,Aug}^T(\beta) \\ &= \frac{1}{n} \sum_{i=1}^n \left(\frac{\delta_i}{\hat{\pi}(v_i)} S(v_i, \beta) + \left(1 - \frac{\delta_i}{\hat{\pi}(v_i)}\right) \hat{m}(z_i, \gamma) \right)^2 \\ &= \frac{1}{n} \sum_{i=1}^n \left(\frac{\delta_i}{\hat{\pi}(v_i)} S(v_i, \beta) + \left(1 - \frac{\delta_i}{\hat{\pi}(v_i)}\right) m(z_i, \gamma) \right)^2 + o_P(1) \\ &= \frac{1}{n} \sum_{i=1}^n \left(\frac{\delta_i}{\pi(v_i)} S(v_i, \beta) + \left(1 - \frac{\delta_i}{\pi(v_i)}\right) m(z_i, \gamma) \right)^2 \\ &\quad + \frac{1}{n} \sum_{i=1}^n \left(\left(\frac{\delta_i}{\hat{\pi}(v_i)} S(v_i, \beta) + \left(1 - \frac{\delta_i}{\hat{\pi}(v_i)}\right) m(z_i, \gamma) \right)^2 \right. \\ &\quad \left. - \left(\frac{\delta_i}{\pi(v_i)} S(v_i, \beta) + \left(1 - \frac{\delta_i}{\pi(v_i)}\right) m(z_i, \gamma) \right)^2 \right) + o_P(1) \\ &= J_1^* + J_2^* + o_P(1), \end{aligned}$$

where J_2^* is equivalent to

$$J_2^* = \frac{1}{n} \sum_{i=1}^n \left(\frac{\delta_i}{\hat{\pi}^2(v_i)} - \frac{\delta_i}{\pi^2(v_i)} \right) S^2(v_i, \beta)$$

$$\begin{aligned}
& + \left(\left(1 - \frac{\delta_i}{\hat{\pi}(v_i)} \right)^2 - \left(1 - \frac{\delta_i}{\pi(v_i)} \right)^2 \right) m(z_i, \gamma)^2 \\
& + \left(\frac{\delta_i}{\hat{\pi}(v_i)} \left(1 - \frac{\delta_i}{\hat{\pi}(v_i)} \right) - \frac{\delta_i}{\pi(v_i)} \left(1 - \frac{\delta_i}{\pi(v_i)} \right) \right) S(v_i, \beta) m(z_i, \gamma) \\
& = J_{21}^* + J_{22}^* + J_{23}^*.
\end{aligned}$$

It is easy to prove that, $\frac{1}{n} \sum_{i=1}^n |S^2(v_i, \beta)| = E(S^2(V, \beta)) + o_P(1)$, $\frac{1}{n} \sum_{i=1}^n |m^2(z_i, \beta)| = E(m^2(Z, \beta)) + o_P(1)$ and $\frac{1}{n} \sum_{i=1}^n |S(v_i, \beta)m(z_i, \beta)| = E(|S(V, \beta)m(Z, \beta)|) + o_P(1)$. Furthermore, by the regularity condition (1) and consistency of $\hat{\pi}$ (to prove that $\hat{\pi} = \pi + o_P(1)$, it is enough to prove consistency of $\hat{\alpha}$ because based on the Eq. (10), $\pi(v) = (1 + \hat{\alpha}(z, \gamma)e^{\gamma c(u)})^{-1}$). By applying SLLN to $\hat{\alpha}$ in Eq. (24), Based on the regularity conditions (2), (3) and (6), we can conclude that $\alpha(z, \gamma) = \frac{E(\delta O(Z, U)|z)}{E(\delta e^{\gamma c(U)}|z)} + o_P(1)$. This with Eq. (8) gives the constancy of $\hat{\alpha}(\cdot)$, we can conclude that $\sup \left| \frac{\delta_i}{\hat{\pi}^2(v_i)} - \frac{\delta_i}{\pi^2(v_i)} \right| = o_P(1)$, $\sup \left| \left(1 - \frac{\delta_i}{\hat{\pi}(v_i)} \right)^2 - \left(1 - \frac{\delta_i}{\pi(v_i)} \right)^2 \right| = o_P(1)$ and $\sup \left| \frac{\delta_i}{\hat{\pi}(v_i)} \left(1 - \frac{\delta_i}{\hat{\pi}(v_i)} \right) - \frac{\delta_i}{\pi(v_i)} \left(1 - \frac{\delta_i}{\pi(v_i)} \right) \right| = o_P(1)$. Therefore, we can write:

$$\begin{aligned}
|J_2^*| & \leq |J_{21}^*| + |J_{22}^*| + |J_{23}^*| \\
& \leq \frac{1}{n} \sum_{i=1}^n |S^2(v_i, \beta)| \times \sup \left(\left| \frac{\delta_i}{\hat{\pi}^2(v_i)} - \frac{\delta_i}{\pi^2(v_i)} \right| \right) \\
& \quad + |m^2(z_i, \beta)| \times \sup \left| \left(1 - \frac{\delta_i}{\hat{\pi}(v_i)} \right)^2 - \left(1 - \frac{\delta_i}{\pi(v_i)} \right)^2 \right| \\
& \quad + |S(v_i, \beta)m(z_i, \beta)| \times \sup \left| \frac{\delta_i}{\hat{\pi}(v_i)} \left(1 - \frac{\delta_i}{\hat{\pi}(v_i)} \right) - \frac{\delta_i}{\pi(v_i)} \left(1 - \frac{\delta_i}{\pi(v_i)} \right) \right|.
\end{aligned}$$

Thus, by the above notations, we can conclude that $J_2^* = o_P(1)$. Therefore, we will have:

$$\frac{1}{n} \sum_{i=1}^n \psi_{i, Aug}(\beta) \psi_{i, Aug}^T(\beta) = \frac{1}{n} \sum_{i=1}^n \left(\frac{\delta_i}{\pi(v_i)} S(v_i, \beta) + \left(1 - \frac{\delta_i}{\pi(v_i)} \right) m(z_i, \gamma) \right)^2$$

The above equation by applying SLLN gives the right hand side of Lemma 1(b). As mentioned, proof of part (c) is very similar to the proof of part (b), therefore we ignore it. Also, for part (a) in a similar way to part (b), we can write:

$$\begin{aligned}\frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{i,IPW}(\beta) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\delta_i}{\hat{\pi}_i(v_i)} S(v_i, \beta) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\delta_i}{\pi_i(v_i)} S(v_i, \beta) + \left(\frac{\delta_i}{\pi_i(v_i)} - \frac{\delta_i}{\hat{\pi}_i(v_i)} \right) S(v_i, \beta).\end{aligned}$$

By applying the same operations used for J_2 to the second summation of right hand side of the above equality, we can write:

$$\begin{aligned}\frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{i,IPW}(\beta) &= \sum_{i=1}^n \frac{\delta_i}{\pi_i(v_i)} S(v_i, \beta) \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^n ((1 - \delta_i) - \delta_i O(v_j)) E(S(V, \beta) | z_i, \delta = 0) \\ &\quad + o_P(1) = o_P(1).\end{aligned}$$

By replacing equivalent functions of $O(v_i)$ and $E(S(V, \beta) | z_i, \delta = 0)$, and some simple mathematical operations we will have:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{i,IPW}(\beta) = \sum_{i=1}^n \frac{\delta_i}{\pi_i(v_i)} S(v_i, \beta) + \left(1 - \frac{\delta_i}{\pi_i(v_i)} \right) m(z_i, \beta) + o_P(1).$$

Therefore, by the above equation and CLT, we can conclude the left hand side of Lemma 1(a). On the other hand, we have:

$$\begin{aligned}\frac{1}{n} \sum_{i=1}^n \psi_{i,IPW}(\beta) \psi_{i,IPW}^T(\beta) &= \frac{1}{n} \sum_{i=1}^n \left(\frac{\delta_i}{\hat{\pi}_i(v_i)} S(v_i, \beta) \right)^2 \\ &= \frac{1}{n} \sum_{i=1}^n \left(\frac{\delta_i}{\pi_i(v_i)} S(v_i, \beta) \right)^2 \\ &\quad + \frac{1}{n} \sum_{i=1}^n \left(\frac{\delta_i}{\pi_i(v_i)} - \frac{\delta_i}{\hat{\pi}_i(v_i)} \right) S(v_i, \beta)^2\end{aligned}$$

$S(\cdot)$ and $\pi(\cdot)$ are bounded and $\hat{\pi} - \pi = o_P(1)$. Therefore, we can conclude that second summation of the above equality is equal to $o_P(1)$. Therefore, we can write:

$$\frac{1}{n} \sum_{i=1}^n \psi_{i,IPW}(\beta) \psi_{i,IPW}^T(\beta) = \frac{1}{n} \sum_{i=1}^n \left(\frac{\delta_i}{\pi_i(v_i)} S(v_i, \beta) \right)^2 + o_P(1).$$

Now by applying SLLN to right hand side of the above equation, we can conclude the right hand side of Lemma 1(a). $\square$

Lemma 2 Under the regularity conditions, for estimated values of $\hat{\gamma}$ from validation sample and true parameter β , we have

$$(a) \quad \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{i,IPW}(\beta) \xrightarrow{D} N(0, \tilde{A}), \quad \frac{1}{n} \sum_{i=1}^n \psi_{i,IPW}(\beta) \psi_{i,IPW}^T(\beta) \xrightarrow{D} B, \quad (32)$$

$$(b) \quad \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{i,Aug}(\beta) \xrightarrow{D} N(0, \tilde{A}), \quad \frac{1}{n} \sum_{i=1}^n \psi_{i,Aug}(\beta) \psi_{i,Aug}^T(\beta) \xrightarrow{D} A, \quad (33)$$

$$(c) \quad \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{i,IA}(\beta) \xrightarrow{D} N(0, \tilde{A}), \quad \frac{1}{n} \sum_{i=1}^n \psi_{i,IA}(\beta) \psi_{i,IA}^T(\beta) \xrightarrow{D} A. \quad (34)$$

Proof Also in this case, we only prove the part (b). Therefore, for Augmented method we have:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{i,Aug}(\beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\delta_i}{\hat{\pi}_i(v_i, \hat{\gamma})} S(v_i, \beta) + \left(1 - \frac{\delta_i}{\hat{\pi}_i(v_i, \hat{\gamma})}\right) \hat{m}(z_i, \hat{\gamma}),$$

define,

$$J_i(\gamma) = \frac{\delta_i}{\hat{\pi}_i(v_i, \gamma)} S(v_i, \beta) + \left(1 - \frac{\delta_i}{\hat{\pi}_i(v_i, \gamma)}\right) \hat{m}(z_i, \gamma) \quad i = 1, \dots, n,$$

Therefore we will have:

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{i,Aug}(\beta) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n J_i(\hat{\gamma}) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (J_i(\gamma) + (J_i(\hat{\gamma}) - J_i(\gamma))) \\ &= J_1 + J_2 + \frac{1}{n} \sum_{i=1}^n \frac{\partial J_i(\gamma)}{\partial \gamma} \Big|_{\gamma=\gamma_0} \sqrt{n}(\hat{\gamma} - \gamma). \end{aligned} \quad (35)$$

In the above equation, J_1 and J_2 are defined the same as that of proof of Lemma 1 and the last statement is concluded from the Intermediate Value Theorem. On the other hand, we have,

$$\begin{aligned} \frac{\partial J_i(\gamma)}{\partial \gamma} &= \delta_i \left(1 - \frac{1}{\hat{\pi}(v_i, \gamma)}\right) (c(u_i) - \hat{m}^*(z_i, \gamma))(S(v_i, \beta) - \hat{m}(z_i, \gamma)) \\ &\quad + \left(1 - \frac{\delta_i}{\hat{\pi}(v_i, \gamma)}\right) (E(S(V, \beta)c(U)|z_i, \delta = 0) - \hat{m}^*(z_i, \gamma)\hat{m}(z_i, \gamma)). \end{aligned}$$

By SLLN we will have:

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \frac{\partial J_i(\gamma)}{\partial \gamma} \\ &= E \left(\delta \left(\frac{1}{\pi(V, \gamma)} - 1 \right) (U - m^*(Z, \gamma))(S(V, \beta) - m(Z, \gamma)) \right) + o_P(1) \\ &= E((1 - \pi(V, \gamma))(U - m^*(Z, \gamma))(S(V, \beta) - m(Z, \gamma))) + o_P(1) \\ &= H + o_P(1). \end{aligned}$$

On the other hand, for $\sqrt{n}(\hat{\gamma} - \gamma)$, we will have:

$$\sum_{i=1}^n r_i(1 - \delta_i)(c(u_i) - \hat{m}^*(z_i, \hat{\gamma})) = 0.$$

By rewriting and using the Intermediate Value Theorem, we obtain that

$$\sum_{i=1}^n r_i(1 - \delta_i)(c(u_i) - \hat{m}^*(z_i, \gamma_0)) - \sum_{i=1}^n r_i(1 - \delta_i) \frac{\partial m^*(z_i, \gamma)}{\partial \gamma} \Big|_{\gamma=\gamma_0} (\hat{\gamma} - \gamma) = 0,$$

where by some mathematical operations, we can conclude from the above equation that:

$$\begin{aligned} \sqrt{n}(\hat{\gamma} - \gamma) &= \left\{ \frac{1}{n} \sum_{i=1}^n r_i(1 - \delta_i) \frac{\partial m^*(z_i, \gamma)}{\partial \gamma} \Big|_{\gamma=\gamma_0} \right\}^{-1} \\ &\quad \times \frac{1}{\sqrt{n}} \sum_{i=1}^n r_i(1 - \delta_i)(c(u_i) - \hat{m}^*(z_i, \gamma_0)) = Q_1^{-1} Q_2, \quad (36) \end{aligned}$$

where,

$$\begin{aligned} Q_1 &= \frac{1}{n} \sum_{i=1}^n r_i(1 - \delta_i)(\hat{E}(c^2(U)|z_i, \delta = 0) - \hat{m}^{*2}(z_i, \gamma)) \\ &= E(r(1 - \delta)(E(c^2(U)|z, \delta = 0) - m^{*2}(Z, \gamma))) + o_P(1) = M + o_P(1). \end{aligned}$$

For Q_2 , we have

$$\begin{aligned} Q_2 &= \frac{1}{\sqrt{n}} \sum_{i=1}^n r_i(1 - \delta_i)(c(u_i) - m^*(z_i, \gamma)) + r_i(1 - \delta_i)(m^*(z_i, \gamma) - \hat{m}^*(z_i, \gamma)) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n r_i(1 - \delta_i)(u_i - m^*(z_i, \gamma)) \end{aligned}$$

$$+ \frac{1}{\sqrt{n}} \sum_{i=1}^n r_i (1 - \delta_i) (m^*(z_i, \gamma) - \hat{m}^*(z_i, \gamma)) = Q_{21} + Q_{22}.$$

For Q_{22} , it is easy to prove that

$$\begin{aligned} Q_{22} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n r_i (1 - \delta_i) \frac{\sum_{j=1}^n \delta_j e^{\gamma c(u_j)} (m^*(z_i, \gamma) - c(u_j)) K_h(z_i, z_j)}{\sum_{j=1}^n \delta_j e^{\gamma c(u_j)} K_h(z_i, z_j)} \\ &= -\frac{1}{n^{\frac{3}{2}}} \sum_{i=1}^n \sum_{j=1}^n r_i (1 - \delta_i) \frac{\delta_j e^{\gamma c(u_j)} (m^*(z_i, \gamma) - c(u_j)) K_h(z_i, z_j)}{\alpha^{-1}(z_i) f(z_i) E(1 - \delta | z_i, \delta = 0)} + o_P(1). \end{aligned}$$

By applying SLLN and some mathematical operations for inner summation of the above equation, we can conclude that:

$$\begin{aligned} Q_{22} &= -\frac{1}{\sqrt{n}} \sum_{j=1}^n \delta_j e^{\gamma c(u_j)} E \left(\frac{r(1 - \delta)(c(U) - m^*(Z, \gamma)) K_h(Z, z_j)}{\alpha^{-1}(z_i) f(z_i) E(1 - \delta | z_i, \delta = 0)} \right) + o_P(1) \\ &= -\frac{1}{\sqrt{n}} \sum_{j=1}^n \delta_j E(r | \delta = 0) \left(\frac{1}{\pi(v_j)} - 1 \right) (u_j - m^*(z_j, \gamma)) + o_P(1). \end{aligned}$$

Therefore,

$$\begin{aligned} Q_2 &= Q_{21} + Q_{22} = \frac{1}{\sqrt{n}} \sum_{i=1}^n (r_i (1 - \delta_i) - \delta_i E(r | \delta = 0)) \\ &\quad \times \left(\frac{1}{\pi(v_i)} - 1 \right) (c(u_i) - m^*(z_i, \gamma)) + o_P(1). \quad (37) \end{aligned}$$

By combining Q_1 and Q_2 in Eq. (37), it is concluded that

$$\begin{aligned} \sqrt{n}(\hat{\gamma} - \gamma) &= M^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n (r_i (1 - \delta_i) - \delta_i E(r | \delta = 0)) \\ &\quad \times \left(\frac{1}{\pi(v_i)} - 1 \right) (c(u_i) - m^*(z_i, \gamma)) + o_P(1). \quad (38) \end{aligned}$$

Finally by inserting Eq. (39) in Eq. (36) and using the obtained details from Eq. (36), we have:

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{i, Aug}(\beta) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \frac{\delta_i}{\pi(v_i)} S(v_i, \beta) + \left(1 - \frac{\delta_i}{\pi(v_i, \gamma)} \right) m(z_i, \gamma) \right. \\ &\quad \left. + H M^{-1} \left(r_i (1 - \delta_i) - \delta_i E(r | \delta = 0) \left(\frac{1}{\pi(v_i)} - 1 \right) \right) \right\} \end{aligned}$$

$$\times (c(u_i) - m^*(z_i, \gamma)).$$

Therefore, the left hand side of Lemma 2(b) can be concluded by applying CLT. Where $\tilde{A}$ is the variance of the following variable,

$$\begin{aligned} \eta &= \frac{\delta}{\pi(V)} S(V, \beta) + \left(1 - \frac{\delta}{\pi(V)}\right) m(Z, \gamma) + H M^{-1} \\ &\times \left(r(1 - \delta) - \delta E(r|\delta = 0) \left(\frac{1}{\pi(V)} - 1\right)\right) (c(U) - m^*(Z, \gamma)). \end{aligned} \quad (39)$$

In a similar way to the proof of right hand side of Lemma 1(b), we can show,

$$\frac{1}{n} \sum_{i=1}^n \psi_{i,Aug}(\beta) \psi_{i,Aug}^T(\beta) = \frac{1}{n} \sum_{i=1}^n \left(\frac{\delta_i}{\pi(v_i)} S(v_i, \beta) + \left(1 - \frac{\delta_i}{\pi(v_i)}\right) m(z_i, \gamma) \right)^2$$

Therefore, By the above equation and SLLN, we can conclude the right hand side of Lemma 2(b). Also, The proof of part (c) is very similar to part (b), therefore we ignore it. Moreover, part (a) obtains in a similar way to proof of Lemma 1(a) and for simplicity, we ignore its details. $\square$

Proof of Theorem 1

For the augmented method, we have:

$$\begin{cases} T_{1,Aug} = \frac{1}{n} \sum_{i=1}^n \frac{\sum_{j=1}^n W_{ij,0} S(z_j, u_j, \beta)}{1 + \sum_{j=1}^n W_{ij,0} \lambda^T S(z_j, u_j, \beta)} = 0 \\ T_{2,Aug} = \frac{1}{n} \sum_{i=1}^n \frac{\sum_{j=1}^n W_{ij,0} \lambda^T h(x_j) h^T(x_j)}{1 + \sum_{j=1}^n W_{ij,0} \lambda^T S(z_j, u_j, \beta)} = 0. \end{cases} \quad (40)$$

Taylor expansion of above equations at $(\beta, 0)$ are as follow:

$$\begin{cases} 0 = T_{1,Aug} = T_{1,Aug}(\beta, 0) + \frac{\partial T_{1,Aug}(\beta, 0)}{\partial \beta} (\hat{\beta} - \beta) + \frac{\partial T_{1,Aug}(\beta, 0)}{\partial \lambda} (\hat{\lambda} - 0) + o_P(\tau_n) \\ 0 = T_{2,Aug} = T_{2,Aug}(\beta, 0) + \frac{\partial T_{2,Aug}(\beta, 0)}{\partial \beta} (\hat{\beta} - \beta) + \frac{\partial T_{2,Aug}(\beta, 0)}{\partial \lambda} (\hat{\lambda} - 0) + o_P(\tau_n), \end{cases} \quad (41)$$

where in Eq. (42) after obtaining derivatives with respect to $(\hat{\beta}, \hat{\lambda})$, they are evaluated in $(\beta, 0)$. Also $\tau_n = \|\hat{\beta} - \beta\| + \|\hat{\lambda}\|$. Note that $T_{1,Aug}(\beta, 0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{j=1}^n W_{ij,0} S(z_j, u_j, \beta)$ and $T_{2,Aug}(\beta, 0) = 0$, so we can conclude from last equation that:

$$\begin{bmatrix} \hat{\lambda} \\ \hat{\beta} - \beta \end{bmatrix} = D^{-1} \begin{bmatrix} -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n W_{ij,0} S(z_j, u_j, \beta) + o_P(\tau_n) \\ o_P(\tau_n) \end{bmatrix}, \quad (42)$$

where in Eq. (43),

$$D = \left[\begin{array}{cc} \frac{\partial T_{1,Aug}(\beta, 0)}{\partial \lambda} & \frac{\partial T_{1,Aug}(\beta, 0)}{\partial \beta} \\ \frac{\partial T_{2,Aug}(\lambda, 0)}{\partial \lambda} & \frac{\partial T_{2,Aug}(\beta, 0)}{\partial \beta} \end{array} \right] \Big|_{(\beta, \lambda) = (\beta, 0)}$$

$$= \left[\begin{array}{cc} -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \psi_{i,Aug}(\beta) \psi_{i,Aug}^T(\beta) & \frac{1}{n} \sum_{i=1}^n \frac{\partial \psi_{i,Aug}(\beta)}{\partial \beta} \\ \frac{1}{n} \sum_{i=1}^n \frac{\partial \psi_{i,Aug}(\beta)}{\partial \beta} & 0 \end{array} \right], \quad (43)$$

where, $\frac{1}{n} \sum_{i=1}^n \frac{\partial \psi_{i,Aug}(\beta)}{\partial \beta} = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{\pi(z_i, \gamma)} h(x_i) h^T(x_i) + (1 - \frac{\delta_i}{\pi(z_i, \gamma)}) E(h(X) h^T(X) | z_i, \delta = 0)$. Therefore,

$$D^{-1} = - \frac{1}{\left(\frac{1}{n} \sum_{i=1}^n \frac{\partial \psi_{i,Aug}(\beta)}{\partial \beta} \right)^2}$$

$$\times \left[\begin{array}{cc} 0 & -\frac{1}{n} \sum_{i=1}^n \frac{\partial \psi_{i,Aug}(\beta)}{\partial \beta} \\ -\frac{1}{n} \sum_{i=1}^n \frac{\partial \psi_{i,Aug}(\beta)}{\partial \beta} & \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \psi_{i,Aug}(\beta) \psi_{i,Aug}^T(\beta) \end{array} \right]. \quad (44)$$

Now, by combining Eqs. (43) and (45) and some simple mathematical operations, we can conclude that:

$$\sqrt{n}(\hat{\beta}_{Aug} - \beta) = \left(\frac{1}{n} \sum_{i=1}^n \frac{\partial \psi_{i,Aug}(\beta)}{\partial \beta} \right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{i,Aug}(\beta) \right) + o_P(1)$$

$$= \left(\frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{\pi(z_i, \gamma)} h(x_i) h^T(x_i) + \left(1 - \frac{\delta_i}{\pi(z_i, \gamma)} \right) E(h(X) h^T(X) | z_i, \delta = 0) \right)^{-1}$$

$$\times \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{i,Aug}(\beta) + o_P(1). \quad (45)$$

Equation (46) and Lemma 1(b) by using SLLN give the following result:

$$\sqrt{n}(\hat{\beta}_{Aug} - \beta) \xrightarrow{D} N(0, T^{-1} A T^{-1}). \quad (46)$$

This proves the second part of Theorem 1(a) for the augmented method. To prove second part of Theorem 1(b), consider Taylor expansion of second order for the empirical log-likelihood ratio of augmented method with respect to parameter λ at 0, i.e.,

$$\hat{l}_{Aug}(\lambda) = \hat{l}_{Aug}(\lambda)|_{\lambda=0} + (\lambda-0)^T \frac{\partial \hat{l}_{Aug}(\lambda)}{\partial \lambda} \Big|_{\lambda=0} + (\lambda-0)^T \frac{\partial^2 \hat{l}_{Aug}(\lambda)}{2 \partial \lambda^2} \Big|_{\lambda=0} (\lambda-0) + o(1), \quad (47)$$

where, $\frac{\partial \hat{l}_{Aug}(\lambda)}{\partial \lambda} = 2 \sum_{i=1}^n \frac{\psi_{Aug}(\beta)}{(1 + \lambda^T \psi_{Aug}(\beta))}$ and $\frac{\partial^2 \hat{l}_{Aug}(\lambda)}{\partial \lambda^2} = -2 \sum_{i=1}^n \frac{\psi_{Aug}(\beta) \psi_{Aug}^T(\beta)}{(1 + \lambda^T \psi_{Aug}(\beta))^2}$. Therefore, the empirical log-likelihood ratio will be as follow:

$$\hat{l}_{Aug}(\lambda) = 2\lambda^T \sum_{i=1}^n \psi_{i,Aug}(\beta) - \lambda^T \sum_{i=1}^n \psi_{i,Aug}(\beta) \psi_{i,Aug}^T(\beta) \lambda + o_P(1), \quad (48)$$

$\hat{\lambda}$ is the solution of $\sum_{i=1}^n \frac{\psi_{i,Aug}(\beta)}{(1 + \lambda^T \psi_{i,Aug}(\beta))} = 0$, thus by expanding of this equation we can get $\hat{\lambda}$ as follow:

$$0 = \sum_{i=1}^n \frac{\psi_{i,Aug}(\beta)}{(1 + \lambda^T \psi_{i,Aug}(\beta))} = \sum_{i=1}^n \psi_{i,Aug}(\beta) + \lambda^T \sum_{i=1}^n \psi_{i,Aug}(\beta) \psi_{i,Aug}^T(\beta) \lambda + o_P(1). \quad (49)$$

Therefore for enough large sample size, we will have:

$$\hat{\lambda} = \left(\sum_{i=1}^n \psi_{i,Aug}(\beta) \psi_{i,Aug}^T(\beta) \right)^{-1} \left(\sum_{i=1}^n \psi_{i,Aug}(\beta) \right) + o_P(1). \quad (50)$$

Finally by combining Eq. (49) with Eq. (51), we conclude that

$$\hat{l}_{Aug} = \left(\sum_{i=1}^n \psi_{i,Aug}(\beta) \right)^T \left(\sum_{i=1}^n \psi_{i,Aug}(\beta) \psi_{i,Aug}^T(\beta) \right)^{-1} \left(\sum_{i=1}^n \psi_{i,Aug}(\beta) \right) + o_P(1). \quad (51)$$

Lemma 1(b) and the relation between normal distribution and chi-square distribution together with the above equation give the second part of Theorem 1(b). The proof of third parts of Theorem 1(a) and (b) are very similar to the proof of second parts of this theorem, therefore we ignore it. By the similar arguments used in the proof of the first part of Theorem 1(a), we can write following equalities in IPW case:

$$\begin{aligned} \sqrt{n}(\hat{\beta}_{IPW} - \beta) &= \left(\frac{1}{n} \sum_{i=1}^n \frac{\partial \psi_{i,IPW}(\beta)}{\partial \beta} \right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{i,IPW}(\beta) \right) + o_P(1) \\ &= \left(\frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{\pi(z_i, \gamma)} h(x_i) h^T(x_i) \right. \\ &\quad \left. + \left(1 - \frac{\delta_i}{\pi(z_i, \gamma)} E(h(X) h^T(X) | z_i, \delta = 0) \right) \right)^{-1} \\ &\quad \times \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{i,IPW}(\beta) + o_P(1). \end{aligned} \quad (52)$$

The above equation and Lemma 1(a) with CLT give the first part of Theorem 1(a). Moreover, In a similar way, we can write:

$$\hat{l}_{IPW} = \left(\sum_{i=1}^n \psi_{i,IPW}(\beta) \right)^T \left(\sum_{i=1}^n \psi_{i,IPW}(\beta) \psi_{i,IPW}^T(\beta) \right)^{-1} \left(\sum_{i=1}^n \psi_{i,IPW}(\beta) \right) + o_P(1). \quad (53)$$

Now, the first part of Theorem 1(a) follows by the above equation and Lemma 1(a).

Proof of Theorem 2

Proof of Theorem 2 follows by the same methods used in Theorem 1 by applying Lemma 2 instead of Lemma 1. Therefore, we ignore the details of the proof of Theorem 2. Also, needing to Lemma 2 in proof of Theorem 2 caused to have different asymptotic distributions in comparison to Theorem 1.

References

- Carpenter JR, Kenward MG, Vansteelandt S (2006) A comparison of multiple imputation and doubly robust estimation for analyses with missing data. *J R Stat Soc Ser A* 169(3):571–584
- Creemers A, Aerts M, Hens H, Molenbergh G (2011) A nonparametric approach to weighted estimating equations for regression analysis with missing covariates. *Comput Stat Data Anal* 56:100–113
- Genback M, Stanghellini E, de Luna (2015) Uncertainty intervals for regression parameters with nonignorable missingness in the outcome. *Stat Pap* 56(3):829–847
- Kim JK, Yu CL (2011) A semiparametric estimation of mean functionals with nonignorable missing data. *J Am Stat Assoc* 106:157–165
- Niu C, Guo X, Xu W, Zhu L (2014) Empirical likelihood inference in linear regression with nonignorable missing response. *Computat Stat Data Anal* 79:91–112
- Owen AB (1988) Empirical likelihood ratio confidence intervals for a single function. *Biometrika* 75:237–249
- Owen AB (1990) Empirical likelihood confidence regions. *Ann Stat* 18:90–120
- Rotnitzky A, Robins J, Scharfstein D (1998) Semiparametric regression for repeated outcomes with nonignorable non-response. *J Am Stat Assoc* 93:1321–1339
- Rubin DB (1976) Inference and missing data. *Biometrika* 63:581–592
- Tang NS, Zhao PY (2013) Empirical likelihood-based inference in nonlinear regression models with missing responses at random. *Statistics* 47:1141–1159
- Wang F, Peng H, Shi W (2016) The relationship between air layers and evaporative resistance of male Chinese ethnic clothing. *Appl Ergon* 56:194–202
- Wang S, Wang CY (2001) A note on kernel assisted estimators in missing covariate regression. *Stat Probab Lett* 55:439–449
- Zhao LX, Zhao PY, Tang NS (2013) Empirical likelihood inference for mean functionals with nonignorable missing response data. *Comput Stat Data Anal* 66:101–116
- Zhao W, Zhang R, Liu Y, Liu J (2015) Empirical likelihood based modal regression. *Stat Pap* 56(2):411–430

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.