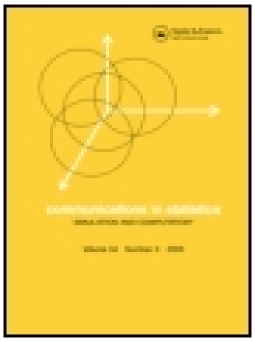

Taylor & Francis
Taylor & Francis Group

Communications in Statistics - Simulation and Computation

ISSN: 0361-0918 (Print) 1532-4141 (Online) Journal homepage: http://www.tandfonline.com/loi/lssp20

A doubly robust goodness-of-fit test in general linear models with missing covariates

Fayyaz Bahari, Safar Parsi & Mojtaba Ganjali

COMMUNICATIONS IN STATISTICS—SIMULATION AND COMPUTATION®
2017, VOL. 0, NO. 0, 1–15
https://doi.org/10.1080/03610918.2016.1255970

Taylor & Francis
Taylor & Francis Group

A doubly robust goodness-of-fit test in general linear models with missing covariates

Fayyaz Bahari[a], Safar Parsi[b], and Mojtaba Ganjali[c]

[a]Department of Statistics, Faculty of Mathematical Sciences, University of Mohaghegh Ardabili, Ardabil, Iran; [b]Department of Engineering Sciences, Faculty of Advanced Technologies, University of Mohaghegh Ardabili, Ardabil, Iran; [c]Department of Statistics, Faculty of Mathematical Sciences, Shahid Beheshti University, Tehran, Iran

ABSTRACT
In this article, we utilize a form of general linear model where missing data occurred randomly on the covariates. We propose a test function based on the doubly robust method to investigate goodness of fit of the model. For this aim, kernel method is used to estimate unknown functions under estimating equation method. Doubly robustness and asymptotic properties of the test function are obtained under local and alternative hypotheses. Furthermore, we investigate the power of the proposed test function by means of some simulation studies and finally we apply this method on analyzing a real dataset.

ARTICLE HISTORY
Received 18 June 2016
Accepted 20 October 2016

KEYWORDS
Doubly robust; Estimating equations; General linear model; Kernel method; Missing data

1. Introduction

We are always seeking ways to describe the relationships among many phenomena. By knowing these relationships, we can predict and program about the future events. Linear model is one of the most common approaches to determine relationships among phenomena. A general linear model represents relation between a response Y and the covariate vector X of dimension m and has the following form:

$$Y = \phi^T(X)\beta + \epsilon, \tag{1}$$

where $\phi(\cdot)$ is a known vector function of dimension p, β is an unknown parameter vector of dimension p, and ϵ is the sampling error with constant variance. Also, it is assumed that $E(\epsilon|X = x) = 0$ and $E(\epsilon^2|X = x) < \infty$. Simple linear model $\phi(X) = X$, is a special case of the general linear model. General linear model is more flexible than simple linear model since it allow us to inference about higher order and more complex models.

In statistical studies, it is necessary to check the validity of specified model to prevent us to give wrong conclusions. In studies without missing values (complete sample studies), Hardle and Mammen (1993) and Hardle et al. (1998) considered testing goodness of fit of the general linear model by nonparametric and semiparametric methods, respectively. Also, when covariates are measured with error, Zhu and Cui (2005) proposed a score-type test function to check goodness of fit of the general linear model.

CONTACT Safar Parsi parsi@uma.ac.ir Department of Engineering Sciences, Faculty of Advanced Technologies, University of Mohaghegh Ardabili, 56199-11367 Ardabil, Iran.

In practice, we always have the problem of missing data, which affect statistical inferences. By using advance statistical and computational tools, some methods are proposed to solve the missing data problem. Dealing with missing data problem has also become easier because of classification of missing data mechanism in three groups by Rubin (1976). In our case, we assume that missing mechanism to be missing at random (MAR) as introduced in Rubin (1976). Furthermore, we assume missing data occur just in some covariates and the response variable is completely observed.

Some authors have assumed that response variable Y to be missing and they have checked goodness of fit of the model. In this case, among the authors, Manteiga and Gonzalez (2006) and Xue (2009) have proposed some test functions based on score-type test function of Hardle et al. (1998). Moreover, Guo and Xu (2012) constructed a test function when some covariates are missing at random. In mentioned literatures usually missed part of data is ignored from estimating equation models and the test functions. Here, we consider estimating models that involve partly missed data and we construct a test function based on partly missed data.

In this article, at first we introduce some methods based on estimating equations method to obtain parameter estimators of a general linear model in the presence of missing data. In the next section, we evaluate our doubly robust test function and we determine its properties. In Section 4, we see the performance of our test statistic based on some simulation studies. In the final section, we use this test function to clarify the goodness of fit of the general linear model by analyzing a real dataset.

2. Methods for estimating parameters of a linear model based on estimating equations method

When we have the complete data, parameters of the general linear model are usually estimated using the following estimating equations:

$$\sum_{i=1}^{n} S_\beta(y_i|x_i) = 0, \tag{2}$$

where $S_\beta(\cdot|\cdot)$ is called score function and it can be the derivative of log-likelihood function, the derivative of sum of squared errors, etc. Using the complete case (CC) method is the simplest method to estimate parameters of a linear model when some parts of covariate X are missing. In this method, missing data completely are discarded. Therefore, partly observed data have no role in the estimation of parameters of the linear model. The parameters' estimates are obtained by solving the following equations:

$$\sum_{i=1}^{n} \delta_i S_\beta(y_i|x_i) = 0, \tag{3}$$

where δ_i is one if the ith individual of data is observed completely and it is zero, otherwise. However, when the missing mechanism is MAR, complete case method gives the bias estimators of the parameters. To solve this problem, Zhao and Lipsitz (1992) proposed inverse probability weight (IPW) method. They showed that estimators of IPW method are unbiased under MAR assumption. In this method, parameters are obtained by solving the following equations:

$$\sum_{i=1}^{n} \frac{\delta_i}{\pi(v_i)} S_\beta(y_i|x_i) = 0, \tag{4}$$

where for the ith individual, $\pi(v_i)$ is the probability of observing ith data conditional on the observed data, that is,

$$\pi(v_i) = P(\delta_i = 1 \mid y_i, x_i) = P(\delta_i \mid y_i, z_i) = P(\delta_i \mid v_i), \tag{5}$$

where $x_i = (u_i, z_i)$ and $v_i = (y_i, z_i)$. The vector z_i includes the fully observed covariates, the vector u_i includes the partly observed covariates, and v_i includes the fully observed dataset. This method gives the unbiased estimators of the parameters but it does not use partly missed data. To solve this problem, Robins et al. (1994, 1995) have proposed the following method to estimate parameters of the linear model:

$$\sum_{i=1}^{n}\left[\frac{\delta_i}{\pi(v_i)}S_\beta(y_i|x_i, z_i) + \left(1 - \frac{\delta_i}{\pi(v_i)}\right)S^*_\beta(v_i)\right] = 0, \tag{6}$$

where $S^*_\beta(v_i) = E(S_\beta(y_i|x_i) \mid V = v_i)$. Since $S^*_\beta(v)$ is the expectation of score function and the inverse probability of observing a datum is used to estimate parameters of the linear model in Eq. (6), this method is called weighted mean score (WMS) method. In WMS method, π and S^* are unknown functions and they need to be estimated. If one of the unknown values of π or S^* specified correctly, then the estimators remain unbiased. Therefore, mean score method is called a doubly robust (DR) method. The unknown values, π and S^* can be estimated in different ways. Wang et al. (1997) and Wang and Wang (2001) have proposed kernel method to estimate these unknown functions. They have estimated these unknown functions by the following estimators:

$$\hat{\pi}(v) = \frac{\sum_{i=1}^{n}\delta_i K_h(v_i - v)}{\sum_{i=1}^{n}K_h(v_i - v)}, \tag{7}$$

$$\hat{S^*}(v_i) = \frac{\sum_{j=1}^{n}\delta_j S_\beta(y_j|x_j)K_h(v_j - v_i)}{\sum_{j=1}^{n}\delta_i K_h(v_j - v_i)}, \tag{8}$$

where $K_h(\cdot)$ is a kernel function with smoothing parameter h. Also, Creemers et al. (2011) have used the following estimator to estimate S^*:

$$\hat{S^*}(v_i) = \frac{\sum_{j=1}^{n}\delta_i S_\beta(y_i|u_j, z_i)K_h(v_j - v_i)}{\sum_{j=1}^{n}\delta_i K_h(v_j - v_i)}, \tag{9}$$

where, in this method, the observed part of the data is used in the estimators. To see the difference between estimators in Eqs. (8) and (9), we refer to Creemers et al. (2011).

For more methods about analyzing linear models with missing covariates, also see Little (1992) and Creemers et al. (2011). Another question that remains to think about after estimating parameters of a general linear model is the validity of the fitted model. To test goodness of fit of the model, Guo and Xu (2012) and Bianco et al. (2013) used IPW method and constructed a test function by using errors. They have considered the following hypotheses:

$$H_0 : Y = \phi^T(X)\beta + \epsilon \quad \text{vs.} \quad H_1 : Y = \phi^T(X)\beta + C_n G(X) + \epsilon, \tag{10}$$

where C_n is a real value constant and $G(\cdot)$ is an unknown function of covariates. In above hypotheses, if $C_n = C$, a constant, the hypothesis is called fixed or global alternative hypothesis; if $C_n \to 0$, as $n \to \infty$, it is called local alternative hypothesis.

To investigate above hypotheses, Guo and Xu (2012) have used IPW method to estimate parameters of the linear model and they have proposed a test function based on errors of this model as follows:

$$T_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i}{\pi(v_i)} (y_i - \phi(x_i)\hat{\beta}). \tag{11}$$

They have shown that T_n tends to a variable with Normal distribution with mean, $\mu_n = -\sqrt{n}C_n(E(G(X)) - E(\phi(X))\Sigma^{-1}E(\phi(X)G(X)))$ and variance, $V_n = E(\frac{1}{\pi(V)}\epsilon^2((1 - E(\phi^T(X))\Sigma^{-1}\phi(X))^2 - (1 - \frac{1}{\pi(V)})((E(\epsilon|V) - E(\phi^T(X))\Sigma^{-1}E(\phi(X)\epsilon|V))^2)$, where $\Sigma = E(\phi^T(X)\phi(X))$. Also, we have modified the variance of their test function by using the same method of them. Their proposed test function have good properties such as $E(T_n) = 0$ under null hypothesis but this test function ignores the partly observed data as does the IPW method in the estimation of linear model parameters. We call their test function as test function based on inverse probability weights (TFIPW). In TFIPW method, we lose partly missed observations. So, to solve this problem and improve the TFIPW method, we propose a test function that uses partly missed data based on WMS method to check the goodness of fit of general linear models.

3. Main results

We proposed a test function based on errors that includes incomplete data (as the WMS method does) in the test function as

$$T = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{\delta_i}{\hat{\pi}(v_i)} e_i + (1 - \frac{\delta_i}{\hat{\pi}(v_i)}) e_i^* \right], \tag{12}$$

where $e_i = y_i - \phi(x_i)\hat{\beta}$ and $e_i^* = y_i - \hat{\phi}^*(v_i)\hat{\beta}$ with $\phi^*(v_i) = E(\phi(X_i)|V_i = v_i)$. Above test function has mean zero if at least one of the two unknown components, $\hat{\pi}(v)$ or $\hat{\phi}^*(v)$ is correctly specified. Therefore, if $\pi(\cdot)$ and $\phi(\cdot)$ are estimated parametrically, T will have the doubly robustness property as WMS method. However, we will estimate them nonparametrically.

In Eq. (12), it is supposed that $\hat{\beta}$ is obtained by WMS method. So in this case, if unknown functions are obtained by similar manner as Eq. (8), we will call proposed test function as test function based on Wang (TFW). On the other hand, if unknown functions are obtained by similar way to that of Eq. (9), we will call proposed test function as test function based on Creemers (TFC). Also in our proposed test functions, $\phi^*(v)$ and $G^*(v) = E(G(X)|V)$ are unknown. Therefore, we will estimate them in a similar way to S^* in Eqs. (8) and (9) for test functions TFW and TFC, respectively.

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

(1) $\pi(v)$ is bounded and has partial derivatives up to order 2 almost surely.
(2) Kernel function $k_h(.)$ is continuous and is from order r. It is always at least from order 2.
(3) The density function of V, $f(v)$, exists and has bounded derivatives up to at least order 2.
(4) The conditional expectations $E(S|V = v)$ and $E(SS^T|V = v)$ 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 converges to infinity. Where d is the dimension of the vector V.

By using following lemma, we can obtain properties of our proposed test function.

Lemma 3.1. *If the regularity conditions hold,*

(a) *Under local hypothesis:*

$$\sqrt{n}(\hat{\beta}-\beta)=\Sigma^{-1}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\frac{\delta_i}{\pi(v_i)}\phi(x_i)\epsilon_i+\left(1-\frac{\delta_i}{\pi(v_i)}\right)\phi^*(v_i)\epsilon_i^*\right]\right\}+O_p(\eta_n) \tag{13}$$

where $\Sigma=E(\phi^T(X)\phi(X))$ *and* $\epsilon^*=E(\epsilon\mid v)$.

(b) *Under alternative hypothesis:*

$$\sqrt{n}(\hat{\beta}-\beta)=\Sigma^{-1}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\frac{\delta_i}{\pi(v_i)}\phi(x_i)\epsilon_i+\left(1-\frac{\delta_i}{\pi(v_i)}\right)\phi^*(v_i)\epsilon_i^*\right]\right.$$
$$\left.+\sqrt{n}C_nE(\phi(X)G(X))\right\}+O_p(\eta_n) \tag{14}$$

Theorem 3.1. *If the regularity conditions hold,*

(a) *Under the local hypothesis:*

$$T\sim N(0,V) \tag{15}$$

where $V=E[\frac{\delta}{\pi(V)}\epsilon(1-E(\phi(X))^T\Sigma^{-1}\phi(X))+(1-\frac{\delta}{\pi(V)})\epsilon^*(1-E(\phi(X))\Sigma^{-1}\phi^*(V))]^2$.

(b) *Under the alternative hypothesis:*

$$T\sim N(\mu,V) \tag{16}$$

where $\mu=-\sqrt{n}C_n[E(G(X))-E(\phi(X))\Sigma^{-1}E(\phi(X)G(X))]$.

The proofs of above lemma and theorem are given in the appendix. Also the estimator of the unknown function can be found among the proofs in the appendix.

4. Simulation study

Some simulations are performed in three studies to investigate the performance of our proposed test function versus TFIPW. To achieve this aim, standard normal kernel function is used to estimate unknown functions and cross-validation method is used to determine bandwidth h. Two different sample sizes, $n=100, 50$, are used to observe the sample size effect on the problem. Any stage of study is repeated 1,000 times and the empirical powers of the test functions are given in Tables 1–3.

Study 1. The data are generated from the following model,

$$Y=\phi^T(X)\beta+C_nG(X)+\epsilon, \tag{17}$$

where $\phi(X)=(1+X^2)$ with $X\sim U(0,1)$, $G(X)=\sin(2\pi X)$, and $\epsilon\sim N(0,0.25)$. As mentioned before we assume that covariate X contains missing data with the following missing at random mechanisms:

Case 1. $\pi_1(y)=1/(1+|y|\exp(-y))$,

Case 2. $\pi_2(y)=1/(1+\exp(-y^2))$.

For above missing mechanisms, it is expected that $E(\pi_1)\simeq 0.78$ and $E(\pi_2)\simeq 0.79$, respectively. Figure 1 shows the curve of Y versus X. Where the dashed curve represents the curve under the null hypothesis and the solid curves represent the curves under the alternative

Table 1. Empirical powers under sample sizes 50 and 100 for different values of C_n in study 1.

	Case 1						Case 2					
	$n=50$			$n=100$			$n=50$			$n=100$		
C_n	TFIPW	TFW	TFC	TFIPW	TFW	TFC	TFIPW	TFW	TFC	TFIPW	TFW	TFC
0.00	0.045	0.066	0.053	0.045	0.066	0.059	0.052	0.091	0.055	0.055	0.061	0.066
0.05	0.053	0.070	0.061	0.064	0.069	0.069	0.074	0.079	0.073	0.085	0.070	0.076
0.10	0.091	0.103	0.099	0.120	0.148	0.130	0.129	0.088	0.115	0.177	0.131	0.157
0.20	0.261	0.273	0.277	0.410	0.441	0.427	0.327	0.213	0.300	0.500	0.396	0.472
0.40	0.676	0.698	0.703	0.935	0.936	0.938	0.745	0.615	0.714	0.938	0.911	0.933
0.60	0.918	0.914	0.927	0.998	0.999	0.999	0.953	0.886	0.954	0.998	0.996	0.999
0.80	0.984	0.981	0.986	0.999	1.000	1.000	0.993	0.965	0.987	1.000	0.999	1.000
1.00	0.999	0.997	0.998	1.000	1.000	1.000	0.998	0.982	0.996	1.000	1.000	1.000
1.20	1.000	0.998	1.000	1.000	1.000	1.000	0.999	0.988	0.999	1.000	1.000	1.000
1.50	1.000	0.998	1.000	1.000	1.000	1.000	1.000	0.995	1.000	1.000	1.000	1.000
2.00	1.000	0.997	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 2. Empirical powers under sample sizes 50 and 100 for different values of C_n in study 2.

	Case 3						Case 4					
	$n=50$			$n=100$			$n=50$			$n=100$		
C_n	TFIPW	TFW	TFC	TFIPW	TFW	TFC	TFIPW	TFW	TFC	TFIPW	TFW	TFC
0.00	0.047	0.107	0.048	0.047	0.070	0.048	0.039	0.209	0.045	0.048	0.173	0.054
0.05	0.056	0.108	0.051	0.046	0.089	0.054	0.041	0.248	0.059	0.054	0.181	0.074
0.10	0.070	0.097	0.058	0.055	0.091	0.068	0.051	0.249	0.064	0.071	0.147	0.077
0.20	0.097	0.120	0.078	0.121	0.112	0.102	0.092	0.229	0.079	0.130	0.171	0.106
0.40	0.175	0.189	0.166	0.300	0.314	0.306	0.171	0.283	0.178	0.291	0.306	0.298
0.60	0.279	0.278	0.288	0.498	0.484	0.500	0.272	0.308	0.295	0.469	0.415	0.482
0.80	0.399	0.413	0.416	0.661	0.649	0.669	0.369	0.378	0.408	0.607	0.527	0.627
1.00	0.475	0.469	0.481	0.773	0.768	0.805	0.440	0.465	0.459	0.713	0.665	0.769
1.20	0.546	0.531	0.578	0.853	0.826	0.855	0.496	0.469	0.564	0.795	0.704	0.822
1.50	0.629	0.584	0.632	0.906	0.875	0.908	0.555	0.488	0.591	0.852	0.749	0.874
2.00	0.690	0.615	0.708	0.943	0.916	0.954	0.592	0.498	0.680	0.907	0.767	0.927

hypothesis. The solid curves keep out from the dashed curve as C_n vary from 0 to 2. Therefore, we can expect that the power of the proposed test function can be increased by increasing C_n.

The empirical powers of the test functions under Study 1 are given in Table 1. From Table 1, it is seen that the power of the test functions converge to 1 as C_n vary from 0 to 1. The power

Table 3. Empirical powers under sample sizes 50 and 100 for different values of C_n in study 3.

	Case 5						Case 6					
	$n=50$			$n=100$			$n=50$			$n=100$		
C_n	TFIPW	TFW	TFC	TFIPW	TFW	TFC	TFIPW	TFW	TFC	TFIPW	TFW	TFC
0.00	0.060	0.193	0.062	0.057	0.148	0.056	0.047	0.233	0.056	0.058	0.164	0.067
0.05	0.058	0.211	0.062	0.052	0.175	0.058	0.047	0.235	0.061	0.057	0.166	0.069
0.10	0.060	0.192	0.065	0.055	0.180	0.060	0.050	0.241	0.060	0.055	0.174	0.072
0.20	0.060	0.221	0.069	0.066	0.229	0.087	0.054	0.258	0.062	0.067	0.202	0.086
0.40	0.082	0.305	0.108	0.137	0.344	0.177	0.079	0.315	0.091	0.103	0.276	0.126
0.60	0.133	0.363	0.163	0.244	0.479	0.300	0.128	0.368	0.137	0.173	0.387	0.205
0.80	0.207	0.431	0.249	0.379	0.608	0.429	0.186	0.418	0.195	0.298	0.484	0.336
1.00	0.302	0.503	0.333	0.495	0.715	0.559	0.252	0.472	0.266	0.407	0.594	0.456
1.20	0.380	0.583	0.427	0.637	0.802	0.693	0.326	0.526	0.350	0.548	0.690	0.591
1.50	0.498	0.662	0.550	0.792	0.903	0.823	0.468	0.605	0.470	0.711	0.796	0.739
2.00	0.700	0.818	0.729	0.930	0.970	0.944	0.616	0.710	0.633	0.881	0.902	0.898

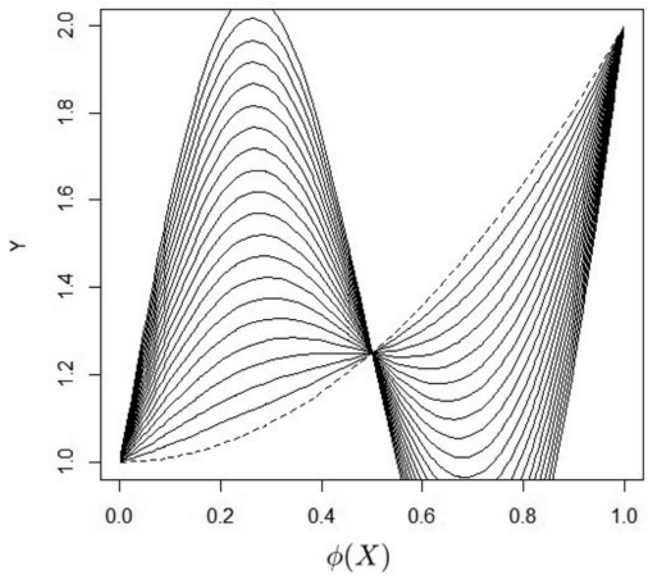

Figure 1. General linear model for different values of C_n. The dashed curve represents the curve under null hypothesis ($C_n = 0$) and the solid curves represent the curves under alternative hypothesis ($C_n = 0.05, 0.10, \ldots, 1$).

of the test functions are affected by sample size. Where the bigger the sample size, the bigger the power of rejection. On the other hand, the power performance has not affected by missing mechanisms. Also from Table 1, we can conclude that all the test functions have approximately the same power of rejection for the same percentage of missing in covariate of the general linear model.

Study 2. In this case, new covariate is added to the linear model. So, the data are generated from the model,

$$Y = \phi^T(X)\beta + C_n G(X) + \epsilon, \tag{18}$$

where $\phi(X) = (1 + X_1^2, X_2)$ with $X_1 \sim U(0, 1)$, $X_2 \sim U(0, 1)$, $G(X) = \sin(2\pi X_2)$, $\beta = (2, 1)$, and $\epsilon \sim N(0, 0.25)$. We assume that missing data occur in X_1 from the following mechanisms:

Case 3. $\pi_3(v) = 1/(1 + |\frac{0.75x_2}{(y+x_2)}|)$,

Case 4. $\pi_4(v) = 1/(1 + |\frac{3x_2}{(y+x_2)}|)$.

In the above cases, the observed full data are in the rates $E(\pi_3(v)) \simeq 0.83$ and $E(\pi_4(v)) \simeq 0.59$, respectively. The results of study 2 are given in Table 2. In case 3 where, missing rate is not high, all the three test functions have approximately the same power of rejection but as C_n becomes bigger, TFC rejects null hypothesis with more probability in comparison with those of the two other test functions. In case 3, when sample size decreases to 50, TFIPW and TFC remain acceptable but TFW rejects null hypothesis near the 0 with the bigger probability. In case 4, where missing rate increases to 41%, TFW shows week conclusions near the $C_n = 0$. On the other hand, TFC shows better performance in comparison with those of the two other test functions. Also by decreasing sample size to 50, the increase in power of TFC is seen more clearly.

Study 3. The data are generated from the following model,

$$Y = \phi^T(X)\beta + C_n G(X) + \epsilon, \tag{19}$$

where $\phi(X) = (1 + X_1^2, X_2, X_3)$ with $X_1 \sim U(0,1), X_2 \sim N(0, 0.25), X_3 \sim U(0,1), G(X) = (X_1 + X_2 + X_3)^2/4$, $\beta = (2, 1, 2)$, and $\epsilon \sim N(0, 0.25)$. In this case, we assume that missing data occur in X_1 and X_2 from the following mechanisms:

Case 5. $\begin{cases} \pi_{51}(v) = 1/(1 + \exp(-(y + x_2 - 0.25x_3))) \\ \pi_{52}(v) = 1/(1 + |0.5y|\exp(-(0.75y + x_1 + x_3))) \end{cases}$

Case 6. $\pi 6 = 1/(1 + 0.5|y + x_3|\exp\{-0.5(y + x_3)\}))$

where in case 5, $1 - \pi_{51}$ is the probability that variable X_1 is missing, while $1 - \pi_{52}$ is the probability that variable X_2 is missed, given that the variable X_1 is observed. Therefore, variables X_1 and X_2 may not be missing at the same time. The probability of missing does not depend on the missed values, therefore, the missing data mechanism is MAR. In case 5, by above missing mechanism, we will have 6 and 4 percentage of missing data in variables X_1 and X_2, respectively. Therefore, in case 5, we will have approximately 90% fully observed data. In case 6, π_6 is the probability of observing data in both variables X_1 and X_2. Therefore, in case 6, both variables X_1 and X_2 are missed or both are observed. By π_6, we expected approximately 80% of fully observed data.

The results of study 3 are given in Table 3. From Table 3, by increasing the number of missed covariates, TFW gives weak results but TFIPW and TFC stay valid. As we expected, TFIPW and TFC have small powers for small values of C_n and have big powers for big values of C_n. From Table 3, we can conclude that TFC is more powerful than TFIPW. Also this fact can be seen for the low sample size ($n = 50$) and the more percentage of missingness (41%) of the case 6.

5. Real data study

We use our proposed test functions to determine fitness of linear model to mono-zygotic twins data in Lee and Scott (1986). Where response Y represents birth weights of a baby, covariate X_1 represents abdominal circumference of a baby, and X_2 represents bi-parietal diameter of a baby. This dataset is also used by Xue (2009) and Guo and Xu (2012) to determine the performance of their methods in linear model with missing data.

Figure 2 represents scatter plot of response Y versus X_1 and X_2 from right to left, respectively. Also we standardized the data, where sample size is 50. From figures, we fit the following linear model:

$$E(Y|X) = \phi^T(X)\beta, \tag{20}$$

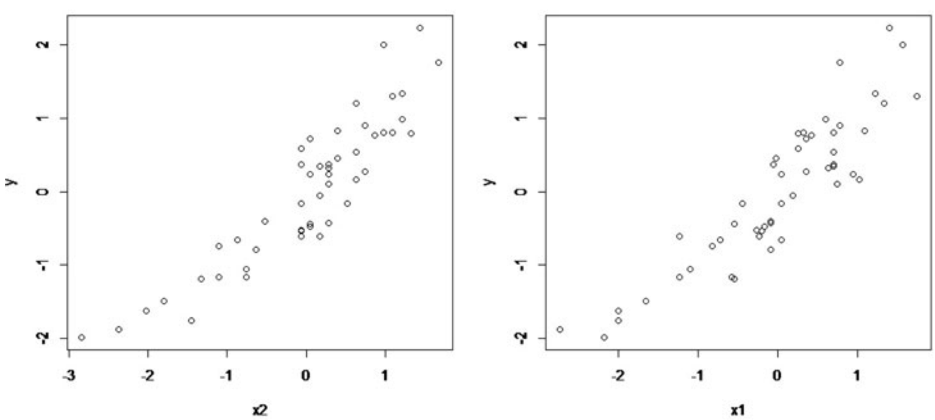

Figure 2. Scatter plot of Y versus X_1 and X_2 from right to left, respectively.

Table 4. *p*-Value for different values of *a*.

a	Missing percentage	TFW	TFC
0.04	5	0.89	0.90
0.07	10	.86	0.87
0.10	13	0.82	0.84
0.11	14	0.82	0.83
0.13	17	0.80	0.81
0.15	18	0.78	0.80
0.18	20	0.76	0.89
0.25	25	0.70	075
0.30	30	0.67	0.73

where $\phi(X) = (X_1, X_2)$ and β is an unknown vector parameter. Also we use the following mechanism for missing data at covariate X_1,

$$\pi(v) = 1/(1 + a|y + x_2|), \tag{21}$$

where a is a known constant to make different percentage of missing. Results of this study is given for different values of a in Table 4. Table 4 represents missing percentage and p-value for TFW and TFC. The p-values give evidence for goodness of fit of the models with different values of a.

6. Discussion

TFIPW has a good performance for low percentage of missingness and its results are acceptable but we should be careful when sample size is not large, missing rate is high, and the number of covariates is high. On the other hand, TFW has a good performance for low rate of missing and large number of data but it has a week performance in other cases. Also the more developed version of it, TFC, has a very good performance for all cases, that is, when the missing rate is high, sample size is large, and the number of covariate is high. For all cases, this method is more powerful than TFIPW.

Appendix: Proof of lemma and theorem

In our proofs, we need asymptotic properties of $\hat{\pi}_i$ and $\hat{\phi}^*_i$. Since $E(\hat{\phi}^*(v_i)) = \phi^*(v_i) + O_p(h^r)$, $\text{Var}(\hat{\phi}^*(v_i)) = O_p(\frac{1}{nh^d})$, $E(\hat{\pi}(v_i)) = \pi(v_i) + O_p(h^r)$, and $\text{Var}(\hat{\pi}(v_i)) = O_p(\frac{1}{nh^d})$, then by Chebyshev inequality we conclude that

$$\hat{\phi}^*(v_i) = \phi^*(v_i) + O_P(\eta_n), \tag{A.1}$$

$$\hat{\pi}(v_i) = \pi(v_i) + O_P(\eta_n). \tag{A.2}$$

Proof of lemma

In proof of lemma, it is enough to proof the part b, because the part (a) is the results of the part (b) by tacking $C_n = 0$. Note that

$$\sqrt{n}(\hat{\beta} - \beta) = \left\{ \frac{1}{n} \sum_{i=1}^{n} \left[\frac{\delta_i}{\hat{\pi}(v_i)} \phi(x_i)\phi^T(x_i) + \left(1 - \frac{\delta_i}{\hat{\pi}(v_i)}\right) \hat{\phi}^*(v_i)\hat{\phi}^{*T}(x_i) \right\}^{-1}$$

$$\cdot\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\frac{\delta_i}{\hat{\pi}(v_i)}\phi(x_i)(y_i-\phi(x_i)\beta)+\left(1-\frac{\delta_i}{\hat{\pi}(v_i)}\right)\hat{\phi}^*(v_i)(y_i-\hat{\phi}^*(v_i)\beta)\right]\right\}=A^{-1}B. \tag{A.3}$$

Since, $\hat{\phi}^*(v_i)=\phi^*(v_i)+O_P(\eta_n)$, we have

$$A=\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\delta_i}{\hat{\pi}(v_i)}\phi(x_i)\phi^T(x_i)+\left(1-\frac{\delta_i}{\hat{\pi}(v_i)}\right)\phi^*(v_i)\phi^{*T}(x_i)+O_P(\eta_n).\right. \tag{A.4}$$

Therefore by expanding sub-function of sigma about $\pi(v_i)$ with respect to $\hat{\pi}(v_i)$ in (A.2), we will have

$$\begin{aligned}A&=\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\delta_i}{\pi(v_i)}\phi(x_i)\phi^T(x_i)+\left(1-\frac{\delta_i}{\pi(v_i)}\right)\phi^*(v_i)\phi^{*T}(x_i)\right]\\&\quad+\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\pi^2(v_i)}(\pi(v_i)-\hat{\pi}(v_i))(\phi(x_i)\phi^T(x_i)-\phi^*(v_i)\phi^{*T}(x_i))+O_P(\eta_n)\\&=A_1+A_2+O_P(\eta_n).\end{aligned} \tag{A.5}$$

By SLLN

$$A_1=E(\phi(X)\phi^T(X))+o_p(1)=\Sigma+o_p(1). \tag{A.6}$$

And for A_2 we have

$$\begin{aligned}A_2&=\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\pi^2(v_i)}(\pi(v_i)-\hat{\pi}(v_i))(\phi(x_i)\phi^T(x_i)-\phi^*(v_i)\phi^{*T}(x_i))+O_P(\eta_n)\\&=\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\pi^2(v_i)}(\phi(x_i)\phi^T(x_i)-\phi^*(v_i)\phi^{*T}(x_i))\frac{\sum_{j=1}^{n}(\pi(v_i)-\delta_j)K_h(v_i-v_j)}{n\hat{f}(v_i)}+O_P(\eta_n).\end{aligned} \tag{A.7}$$

It is easily concluded that

$$\begin{aligned}A_2&=\frac{1}{n}\sum_{i=1}^{n}\frac{\pi(v_i)-\delta_i}{\pi(v_i)}[E(\phi(X)\phi^T(X)\mid z_i)-E(\phi^*(X)\phi^{*T}(X)\mid z_i)]+O_P(\eta_n)\\&=\frac{1}{n}\sum_{i=1}^{n}\frac{\pi(v_i)-\delta_i}{\pi(v_i)}[E(\phi(X)\phi^T(X)\mid z_i)-\phi^*(X)\phi^{*T}(X)]+O_P(\eta_n).\end{aligned} \tag{A.8}$$

For more details about above conclusion see also similar conclusion in Wang and Wang (2001, pp. 447–448). So, from Eq. (A.8) we conclude that

$$A_2=O_p(\eta_n). \tag{A.9}$$

Therefore, from Eqs. (A.5), (A.6), and (A.9) it can be concluded that

$$A=\Sigma+O_p(\eta_n). \tag{A.10}$$

Since, $\hat{\phi}^*(v_i) = \phi^*(v_i) + O_P(\eta_n)$, we have

$$\begin{aligned}
B &= \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{\delta_i}{\hat{\pi}(v_i)} \phi(x_i)(y_i - \phi(x_i)\beta) + \left(1 - \frac{\delta_i}{\hat{\pi}(v_i)}\right) \phi^*(v_i)(y_i - \phi^*(v_i)\beta) \right] \right\} + O_P(\eta_n) \\
&= \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{\delta_i}{\hat{\pi}(v_i)} \phi(x_i)(\epsilon_i + C_n G(x_i)) + \left(1 - \frac{\delta_i}{\hat{\pi}(v_i)}\right) \phi^*(\epsilon_i^* + C_n G^*(v_i)) \right] \right\} + O_P(\eta_n) \\
&= \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{\delta_i}{\hat{\pi}(v_i)} \phi(x_i)\epsilon_i + \left(1 - \frac{\delta_i}{\hat{\pi}(v_i)}\right) \phi^* \epsilon_i^* \right] \right\} \\
&\quad + \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{\delta_i}{\hat{\pi}(v_i)} \phi(x_i) C_n G(x_i) + \left(1 - \frac{\delta_i}{\hat{\pi}(v_i)}\right) \phi^* C_n G^*(v_i) \right] \right\} + O_P(\eta_n). \qquad \text{(A.11)}
\end{aligned}$$

Therefore by expanding sub-function of sigma about $\pi(v_i)$ with respect to $\hat{\pi}(v_i)$, we have

$$\begin{aligned}
B &= \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{\delta_i}{\pi(v_i)} \phi(x_i)\epsilon_i + \left(1 - \frac{\delta_i}{\pi(v_i)}\right) \phi^*(v_i)\epsilon_i^* \right\} \right. \\
&\quad + \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{\delta_i}{\pi(v_i)} \phi(x_i) C_n G(x_i) + \left(1 - \frac{\delta_i}{\pi(v_i)}\right) \phi^*(v_i) C_n G^*(v_i) \right\} \right. \\
&\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{\delta_i}{\pi^2(v_i)} (\phi(x_i)(y_i - \phi(x_i)\beta) - \phi^*(v_i)(y_i - \phi^*(v_i)\beta))(\hat{\pi}(v_i) \right. \\
&\quad - \pi(v_i) + O_P(\eta_n) = B_1 + B_2 + O_P(\eta_n). \qquad \text{(A.12)}
\end{aligned}$$

By SLLN for B_1 we have

$$\begin{aligned}
B_1 &= \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{\delta_i}{\pi(v_i)} \phi(x_i)\epsilon_i + \left(1 - \frac{\delta_i}{\pi(v_i)}\right) \phi^*(v_i)\epsilon_i^* \right\} \right. \\
&\quad + \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{\delta_i}{\pi(v_i)} \phi(x_i) C_n G(x_i) + \left(1 - \frac{\delta_i}{\pi(v_i)}\right) \phi^*(v_i) C_n G^*(x_i \right\} \right. \\
&= \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{\delta_i}{\pi(v_i)} \phi(x_i)\epsilon_i + \left(1 - \frac{\delta_i}{\pi(v_i)}\right) \phi^*(v_i)\epsilon_i^* \right\} + \sqrt{n} C_n E(\phi(X)G(X)) + o_p(1). \right.
\end{aligned}$$

$$\text{(A.13)}$$

For B_2, it can easily be concluded similar to Eq. (A.7) that

$$\begin{aligned}
B_2 &= \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{\delta_i}{\pi^2(v_i)} (\phi(x_i)(y_i - \phi(x_i)\beta) - \phi^*(v_i)(y_i - \phi^*(v_i)\beta)) \right. \right. \\
&\quad \times \frac{\sum_{j=1}^{n} (\phi(x_i) - \delta_j) K_h(v_i - v_j)}{n \hat{f}(v_i)} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\pi(v_i) - \delta_i}{\pi(v_i)} E(\phi(x_i)(y_i - \phi(x_i)\beta) - \phi^*(v_i)(y_i - \phi^*(v_i)\beta) \mid v_i) + O_P(\eta_n)
\end{aligned}$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\pi(v_i) - \delta_i}{\pi(v_i)} [\phi^*(v_i) y_i - E(\phi\phi \mid v_i)\beta - \phi^*(v_i) y_i + \phi^*(v_i)\phi^{*T}(x_i)\beta] + O_P(\eta_n)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\pi(v_i) - \delta_i}{\pi(v_i)} [\phi^*(v_i)\phi^{*T}(x_i) - E(\phi(X)\phi^T(X) \mid v_i)]\beta + O_P(\eta_n) = O_P(\eta_n). \tag{A.14}$$

Therefore from Eqs. (A.13) and (A.14), we can conclude that

$$B = \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{\delta_i}{\pi(v_i)} \phi(x_i)\epsilon_i + \left(1 - \frac{\delta_i}{\pi(v_i)}\right) \phi^*(v_i)\epsilon_i^* \right\} + \sqrt{n} C_n E(\phi(x_i)G(x_i)) + O_P(\eta_n). \tag{A.15}$$

Finally, the part (b) of lemma can be concluded from Eqs. (A.10) and (A.15). As mentioned before, the part (a) can be concluded as the part (b) by tacking $C_n = 0$.

Proof of theorem

The test function is

$$T = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i}{\hat{\pi}(v_i)} (y_i - \phi(x_i)\hat{\beta}) + \left(1 - \frac{\delta_i}{\hat{\pi}(v_i)}\right) (y_i - \hat{\phi}^*(v_i)\hat{\beta}), \tag{A.16}$$

since $\hat{\phi}^*(v_i) = \phi^*(v_i) + O_P(\eta_n)$, then we can write T as

$$T = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i}{\hat{\pi}(v_i)} (y_i - \phi(x_i)\hat{\beta}) + \left(1 - \frac{\delta_i}{\hat{\pi}(v_i)}\right) (y_i - \phi^*(v_i)\hat{\beta}) + O_P(\eta_n). \tag{A.17}$$

Also T can be written as below:

$$T = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i}{\hat{\pi}(v_i)} (y_i - \phi(x_i)\beta) + \left(1 - \frac{\delta_i}{\hat{\pi}(v_i)}\right) (y_i - \phi^*(v_i)\beta)$$
$$- \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i}{\hat{\pi}(v_i)} \phi(x_i)(\beta - \hat{\beta}) + \left(1 - \frac{\delta_i}{\hat{\pi}(v_i)}\right) \phi^*(v_i)(\beta - \hat{\beta}) + O_P(\eta_n). \tag{A.18}$$

Now, we expand T about $\pi(v_i)$, with respect to $\hat{\pi}(v_i)$:

$$T = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{\delta_i}{\pi(v_i)} (y_i - \phi(x_i)\beta) + \left(1 - \frac{\delta_i}{\pi(v_i)}\right) (y_i - \phi^*(v_i)\beta) | \right]$$
$$- \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{\delta_i}{\pi(v_i)} \phi(x_i)(\beta - \hat{\beta}) + \left(1 - \frac{\delta_i}{\pi(v_i)}\right) \phi^*(v_i)(\beta - \hat{\beta}) | \right]$$
$$+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{\delta_i}{\pi^2(v_i)} (y_i - \phi(x_i)\beta) - \frac{\delta_i}{\pi^2(v_i)} (y_i - \phi^*(v_i)\beta)][\pi(v_i) - \hat{\pi}(v_i) | \right]$$
$$- \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{\delta_i}{\pi^2(v_i)} \phi(x_i)(\beta - \hat{\beta}) - \frac{\delta_i}{\pi^2(v_i)} \phi^*(v_i)(\beta - \hat{\beta}) \right] [\pi(v_i) - \hat{\pi}(v_i)] + O_P(\eta_n)$$
$$= T_1 + T_2 + T_3 + T_4 + O_P(\eta_n). \tag{A.19}$$

For T_1 we have

$$
\begin{aligned}
T_1 &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{\delta_i}{\pi(v_i)} (y_i - \phi(x_i)\beta) + \left(1 - \frac{\delta_i}{\pi(v_i)}\right) (y_i - \phi^*(v_i)\beta) \right] \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{\delta_i}{\pi(v_i)} \epsilon_i + \left(1 - \frac{\delta_i}{\pi(v_i)}\right) \epsilon_i^*) \right] \\
&\quad + \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{\delta_i}{\pi(v_i)} C_n G(x_i) + \left(1 - \frac{\delta_i}{\pi(v_i)}\right) C_n G^*(x_i) \right] \right. \\
&= \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{\delta_i}{\pi(v_i)} \epsilon_i + \left(1 - \frac{\delta_i}{\pi(v_i)}\right) \epsilon_i^* \right\} + \sqrt{n} C_n E(G(X)) + o_p(1). \qquad \text{(A.20)}
\end{aligned}
$$

We can rewrite T_2 as follows:

$$
T_2 = -\frac{1}{n} \sum_{i=1}^{n} \left[\frac{\delta_i}{\pi(v_i)} \phi(x_i) + \left(1 - \frac{\delta_i}{\pi(v_i)}\right) \phi^*(v_i) \right] [\sqrt{n}(\beta - \hat{\beta})] \qquad \text{(A.21)}
$$

and by using part (b) of lemma, we can write T_2 as

$$
\begin{aligned}
T_2 = &-\frac{1}{n} \sum_{i=1}^{n} \left[\frac{\delta_i}{\pi(v_i)} \phi(x_i) + \left(1 - \frac{\delta_i}{\pi(v_i)}\right) \phi^*(v_i) \right] \\
&. \left[\Sigma^{-1} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{\delta_i}{\pi(v_i)} \phi(x_i)\epsilon_i + \left(1 - \frac{\delta_i}{\pi(v_i)}\right) \phi^*(v_i)\epsilon_i^* \right] \right. \right. \\
&+ \sqrt{n} C_n E(\phi(X)G(X)) \Big\} \Big] + O_p(\eta_n). \qquad \text{(A.22)}
\end{aligned}
$$

So, we rewrite T_2 as

$$
\begin{aligned}
T_2 = &-E(\phi(X_i))\Sigma^{-1} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{\delta_i}{\pi(v_i)} \phi(x_i)\epsilon_i + \left(1 - \frac{\delta_i}{\pi(v_i)}\right) \phi^*(v_i)\epsilon_i^* \right] \right. \\
&+ \sqrt{n} C_n E(\phi(X)G(X))| \Big\} + O_p(\eta_n). \qquad \text{(A.23)}
\end{aligned}
$$

For T_3, we have

$$
\begin{aligned}
T_3 &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{\delta_i}{\pi^2(v_i)} (y_i - \phi(x_i)\beta) - \frac{\delta_i}{\pi^2(v_i)} (y_i - \phi^*(v_i)\beta)][\pi(v_i) - \hat{\pi}(v_i) \right] \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i}{\pi^2(v_i)} [(\phi^*(v_i) - \phi(x_i))\beta)][\pi(v_i) - \hat{\pi}(v_i)] \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{\delta_i}{\pi^2(v_i)} ([(\phi^*(v_i) - \phi(x_i))\beta)) \frac{\sum_{j=1}^{n} (\pi(v_i) - \delta_j) K_h(v_i - v_j)}{n\hat{f}(v_i)} \right]. \qquad \text{(A.24)}
\end{aligned}
$$

Similar to Eq. (A.7), we can write T_3 as below:

$$T_3 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{\pi(v_i) - \delta_i}{\pi(v_i)} \beta E(\phi^*(v_i) - \phi(x_i) \mid v_i) \right] + O_p(\eta_n)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{\pi(v_i) - \delta_i}{\pi(v_i)} \beta(\phi^*(v_i) - \phi^*(v_i) \right] + O_p(\eta_n) = O_p(\eta_n). \tag{A.25}$$

And we can write T_4 as

$$T_4 = \left[\sqrt{n}(\hat{\beta} - \beta)\right] \frac{1}{n} \sum_{i=1}^{n} \left[\frac{\delta_i}{\pi^2(v_i)} [\phi(x_i) - \phi^*(v_i)][\pi(v_i) - \hat{\pi}(v_i)] \right]$$

$$= \left[\sqrt{n}(\hat{\beta} - \beta)\right] \frac{1}{n} \sum_{i=1}^{n} \left[\frac{\delta_i}{\pi^2(v_i)} [\phi(x_i) - \phi^*(v_i)] \left[\frac{\sum_j^n (\pi(v_i) - \delta_j) K_h(v_i - v_j)}{n \hat{f}(v_i)} \right] \right]. \tag{A.26}$$

Similar to Eq. (A.7), we can write

$$T_4 = \left[\sqrt{n}(\hat{\beta} - \beta)\right] \left\{ \frac{1}{n} \sum_{i=1}^{n} \left[\frac{\pi(v_i) - \delta_i}{\pi(v_i)} [E(\phi(x_i) \mid v_i) - E(\phi^*(v_i) \mid v_i) \right] \right\} + O_p(\eta_n)$$

$$= \left[\sqrt{n}(\hat{\beta} - \beta)\right] \left\{ \frac{1}{n} \sum_{i=1}^{n} \left[\frac{\pi(v_i) - \delta_i}{\pi(v_i)} [\phi^*(v_i) - \phi^*(v_i) \right] \right\} + O_p(\eta_n) = O_p(\eta_n). \tag{A.27}$$

Therefore, by Eqs. (A.20), (A.23), (A.25), and (A.27), we can write T as

$$\begin{aligned} T &= T_1 + T_2 + T_3 + T_4 + O_P(\eta_n) = \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{\delta_i}{\pi(v_i)} \epsilon_i + \left(1 - \frac{\delta_i}{\pi(v_i)}\right) \epsilon_i^* \right\} + \sqrt{n} C_n E(G(X)) \right. \\ &\quad - E(\phi(X_i)) \Sigma^{-1} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{\delta_i}{\pi(v_i)} \phi(x_i) \epsilon_i + \left(1 - \frac{\delta_i}{\pi(v_i)}\right) \phi^*(v_i) \epsilon_i^* \right] \right. \\ &\quad \left. + \sqrt{n} C_n E(\phi(X) G(X)) \right\} + O_p(\eta_n), \end{aligned} \tag{A.28}$$

and we can rewrite T as follows:

$$\begin{aligned} T &= \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{\delta_i}{\pi(v_i)} \epsilon_i (1 - E(\phi(X)) \Sigma^{-1} \phi(x_i)) + \left(1 - \frac{\delta_i}{\pi(v_i)}\right) \epsilon_i^* (1 - E(\phi(X)) \Sigma^{-1} \phi^*(v_i)) \right. \right. \\ &\quad \left. \left. + \sqrt{n} C_n (E(G(X)) - E(\phi(X)) \Sigma^{-1} E(\phi(X) G(X)) \right\} \right) + O_p(\eta_n). \end{aligned} \tag{A.29}$$

Finally the part (b) of theorem can be concluded from Central Limit Theorem as n converge to infinity. The part (a) of theorem can be concluded similar to that of the part (b) by tacking $C_n = 0$.

References

Bianco, A. M., Boente, G., Rodrigues, I. M. (2013). Robust test in generalized linear models with missing responses. *Computational Statistics and Data Analysis* 65:80–97.

Creemers, A., Aerts, M., Hens, H., Molenbergh, G. (2011). A nonparametric approach to weighted estimating equations for regression analysis with missing covariates. *Computational Statistics and Data Analysis* 56:100–113.

Guo, X., Xu, W. (2012). Goodness-of-fit tests for general linear models with covariates missed at random. *Journal of Statistical Planning and Inference* 142:2047–2058.

Hardle, W., Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Annals of Statistics* 21:1921–1947.

Hardle, W., Mammen, E., Muller, M. (1998). Testing parametric versus semiparametric modeling in generalized linear models. *Journal of the American Statistical Association* 93:1461–1474.

Lee, A. J., Scott, A. J. (1986). Ultrasound in ante-natal diagnosis. In: Brook, R. J., Arnold, G. C., Hassard, T. H., Pringle, R. M., eds. *The fascination of Statistics*. New York: Marcel Decker, pp. 277–293.

Little, R. J. A. (1992). Regression with missing X's: A review. *Journal of the American Statistical Association* 87:1227–1237.

Manteiga, G. W., Gonzalez, P. A. (2006). Goodness-of-fit tests for linear regression models with missing response data. *Canadian Journal of Statistics* 34:149–170.

Robins, J., Rotnitzky, A., Zhao, L. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* 89:846–866.

Robins, J., Rotnitzky, A., Zhao, L. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* 90:106–121.

Rubin, D. B. (1976). Inference and missing data. *Biometrika* 63:581–592.

Wang, C. Y., Wang, S., Zhao, L. P., Ou, S. T. (1997). Weighted semiparametric estimation in regression analysis with missing covariate data. *Journal of the American Statistical Association* 92:512–525.

Wang, S., Wang, C. Y. (2001). A note on kernel assisted estimators in missing covariate regression. *Statistics and Probability Letters* 55:439–449.

Xue, L. G. (2009). Empirical likelihood for linear models with missing responses. *Journal of Multivariate Analysis* 100:1353–1366.

Zhao, L. P., Lipsitz, S. (1992). Design and analysis of two-stage studies. *Statistics in Medicine* 11:769–782.

Zhu, L. X., Cui, H. J. (2005). Testing lack-of-fit for general linear errors in variables models. *Statistica Sinica* 15:1049–1068.