Research Article

Empirical Likelihood for Multidimensional Linear Model with Missing Responses

Liping Zhu

Department of Mathematics, Changji College, Changji, Xinjiang 831100, China

Correspondence should be addressed to Liping Zhu, lipingzhu2002@163.com

Imputation is a popular technique for handling missing data especially for plenty of missing values. Usually, the empirical log-likelihood ratio statistic under imputation is asymptotically scaled chi-squared because the imputing data are not i.i.d. Recently, a bias-corrected technique is used to study linear regression model with missing response data, and the resulting empirical likelihood ratio is asymptotically chi-squared. However, it may suffer from the "the curse of high dimension" in multidimensional linear regression models for the nonparametric estimator of selection probability function. In this paper, a parametric selection probability function is introduced to avoid the dimension problem. With the similar bias-corrected method, the proposed empirical likelihood statistic is asymptotically chi-squared when the selection probability is specified correctly and even asymptotically scaled chi-squared when specified incorrectly. In addition, our empirical likelihood estimator is always consistent whether the selection probability is specified correctly or not, and will achieve full efficiency when specified correctly. A simulation study indicates that the proposed method is comparable in terms of coverage probabilities.

1. Introduction

Consider the following multidimensional linear regression model:

$$Y_i = X_i'\beta + \epsilon_i, \quad 1 \le i \le n, \tag{1.1}$$

where Y_i is a scalar response variable, X_i is a $p \times 1$ vector of design variable, β is a $p \times 1$ vector of regression parameter, and the errors ϵ_i are independent random variables with $E[\epsilon_i \mid X_i] = 0$, $\text{Var}[\epsilon_i \mid X_i] = \sigma^2$. Suppose that we have incomplete observations (X_i, Y_i, δ_i), $i = 1, 2, \ldots, n$, from this model, where all the $X_i's$ are observed, and $\delta_i = 0$ if Y_i is missing, $\delta_i = 1$ otherwise. Throughout this paper, we assume that Y is missing at random (MAR); that is, the probability

that Y is missing may depend on X but not on, Y that is, $P(\delta = 1 \mid X, Y) = P(\delta = 1 \mid X)$. We focus on constructing confidence regions on β with the incomplete data (X_i, Y_i, δ_i), $i = 1, 2, \ldots, n$.

The empirical likelihood (EL) method, introduced by Owen [1], has many advantages over normal approximation methods for constructing confidence intervals [1]. One is that it produces confidence intervals and regions whose shape and orientation are determined entirely by the data; the other is that empirical likelihood regions are range preserving and transformation respecting. But it cannot directly to be used in missing responses situation. A natural method is to impute the predictor of Y based on the completely observed pairs, which is provided by Wang and Rao [2]. Unfortunately, the empirical log-likelihood ratio under this imputation is always asymptotically scaled chi-squared. Therefore, the empirical log-likelihood ratio cannot be applied directly to make a statistical inference on the parameter.

Recently, a bias-corrected technique combined imputation method, and Horvitz-Thompson inverse-selection weighted method is separately explored by Xue [3] and Qin et al. [4]. The provided empirical likelihood ratios obey Wilks theorem and the empirical likelihood estimator is consistent. However, the true selection probability function is ordinarily unknown, its nonparametric estimator may suffer from "the curse of high dimensionality" in multidimensional linear regression models.

To avoid "the curse of high dimensionality", it is customary to suppose a parametric selection probability function, but the risk of misspecified function maybe exist. In this paper, we consider the following two situations, one is that the selection probability is specified correctly, the other is that specified incorrectly. To the best of our knowledge, this issue is rarely to be discussed. With the similar bias-corrected method, the provided empirical likelihood statistic is asymptotically chi-squared in the first situation and asymptotically weighted sum of chi-squared in the second situation. In addition, the following desired feature is worth mentioning. The auxiliary random vector to construct the empirical likelihood statistic is just the same form as that of proposed by Robins et al. [5]. The Robins' estimator has the characteristic of "doubly robust", that is, the estimator is asymptotically unbiased either the underlying missing data mechanism or the underlying regression function is correctly specified. Because the underlying regression function is asymptotically correct in our situation, our estimator is always consistent in both cases. Even our estimator can achieve asymptotically full efficiency in the first situation. From this point, it is feasible to use the parametric selection probability function to construct an EL statistic.

The rest of the paper is organized as follows. In Section 2, separately, the EL statistic for β is constructed and the asymptotically results are shown in the two situations. Section 3 reports some simulation results on the performance of the proposed EL confidence region. The proof of the main result is given in the Appendix.

2. Main Result

For simplicity, denote the true selection probability function by $p_0(x)$ and a specified probability distribution function by $p(x, \theta)$ for given θ, a $p \times 1$ unknown vector parameter. Thus, the first situation means $p_0(x) = p(x, \theta_0)$ for some θ_0, where θ_0 is the true parameter and the second situation means $p_0(x) \neq p(x, \theta)$ for any θ. And let $\hat{\beta}_r$ be the least square estimator of β based on the completely observed pairs $(X_i, Y_i, \delta_i = 1)$, $i = 1, 2, \ldots, n$ that is, $\hat{\beta}_r = (\sum_{i=1}^{n} \delta_i X_i X_i')^{-1} \sum_{i=1}^{n} \delta_i X_i Y_i$.

2.1. Empirical Likelihood

In this subsection, an empirical likelihood statistic is conducted, and then some asymptotic results are shown when the selection probability is specified correctly, that is, $p_0(x) = p(x, \theta_0)$ for some θ_0.

Since the design variable X_i is observable for each subject, the maximum likelihood estimator $\hat{\theta}$ can be obtained by maximizing the likelihood function

$$L(\theta) = \prod_{i=1}^{n} p(X_i, \theta)^{\delta_i} \left(1 - p(X_i, \theta)\right)^{1-\delta_i}. \tag{2.1}$$

The following regularity assumptions on $L(\theta)$ are sufficiently strong to ensure both consistency and asymptotic normality of the maximum likelihood estimator. Suppose $U(\theta_0)$ is some field of θ_0 where θ_0 is the true parameter:

(C1) $\partial \ln L_B(\theta)/\partial \theta$, $\partial^2 \ln L_B(\theta)/\partial \theta^2$ and $\partial^3 \ln L_B(\theta)/\partial \theta^3$ exist in $U(\theta_0)$ for all X;

(C2) $|\partial^3 \ln L_B(\theta)/\partial \theta^3| \leq H(X)$ in $U(\theta_0)$, and $EH(X) < \infty$;

(C3) $E_{\theta_0}[\ln L'_B(\theta)/\ln L_B(\theta)] = 0$, $E_{\theta_0}[\ln L''_B(\theta)/\ln L_B(\theta)] = 0$, $I(\theta_0) = E_{\theta_0}[\ln L'_B(\theta)/\ln L_B(\theta)]^2 > 0$.

Then, we use $\check{Y}_i = (\delta_i/p(X_i, \hat{\theta}))Y_i + (1 - \delta_i/p(X_i, \hat{\theta}))X'_i\hat{\beta}_r$, $i = 1, 2, \ldots, n$, as "complete" data set for Y to construct the auxiliary random vectors

$$Z^*_{\text{in}}(\beta) = X_i(\check{Y}_i - X'_i\beta) = X_i \left\{ \frac{\delta_i}{p\left(X_i, \hat{\theta}\right)}(Y_i - X'_i\beta) + \left(1 - \frac{\delta_i}{p\left(X_i, \hat{\theta}\right)}\right)X'_i\left(\hat{\beta}_r - \beta\right) \right\}. \tag{2.2}$$

Thus, an empirical log-likelihood ratio is defined as

$$l^*_n(\beta) = -2\max\left\{ \sum_{i=1}^{n} \log(n\omega_i) \mid \omega_i \geq 0, \sum_{i=1}^{n} \omega_i = 1, \sum_{i=1}^{n} \omega_i Z^*_{\text{in}}(\beta) = 0 \right\}. \tag{2.3}$$

Further, the maximum empirical likelihood estimator $\tilde{\beta}$ of β is to maximize $\{-l^*_n(\beta)\}$.

To ensure asymptotic results, the following assumptions are needed:

(C4) $p(x, \theta)$ is uniformly continued in $U(\theta_0)$ for all X;

(C5) A, D_1 is a positive definite matrix, where $A = E\{XX^T\}$ and $D_1 = E\{\{1/p(X, \theta_0)\}XX'\varepsilon^2\}$;

(C6) $p(x, \theta_0)$ has bounded almost surely and in $f_x p(x, \theta_0) > 0$;

where condition (C4) is common for some selection probability function. Condition, (C5)-(C6) are necessary for asymptotic normality of the maximum empirical likelihood estimator.

Theorem 2.1. *Suppose that conditions (C1)–(C6) above hold. If β is the true parameter and $p(x, \theta)$ is specified correctly, then*

$$l_n^*(\beta) \xrightarrow{D} \chi_p^2, \tag{2.4}$$

where χ_p^2 means the chi-square variable with p degrees of freedom. $\xrightarrow{D}$ represents the convergence in distribution.

Let the $\chi_p^2(1 - \alpha)$ be the $1 - \alpha$ quartile of the χ_p^2 for $0 < \alpha < 1$. Using *Theorem 2.1*, we obtain an approximate $1 - \alpha$ confidence region for β, defined by

$$\left\{ \check{\beta} \mid l_n^*(\check{\beta}) \leq \chi_p^2(1 - \alpha) \right\}. \tag{2.5}$$

Theorem 2.1 can also be used to test the hypothesis $H_0 : \beta = \beta_0$. One could reject H_0 if $l_n^*(\check{\beta}) > \chi_p^2(1 - \alpha)$.

Remark 2.2. In general, plug-in empirical likelihood will asymptotically lead to a sum of weighted χ_1^2 variables with unknown weights for the EL statistics proposed. However, when $p(x, \theta)$ is specified correctly, the EL statistic with two plug-ins has the limiting distribution of χ_p^2 which is due to the following reasons. Firstly, the bias-correction method, that is, the selection function as inverse weight will eliminate the influence by the $\widehat{\beta}_r$. Secondly, the estimating function has special structure, that is, the influence of $\widehat{\theta}$ will be also eliminated if θ is concluded in the denominator of function.

Theorem 2.3. *Under Conditions (C1)–(C6), if $p(x, \theta)$ is specified correctly, then*

$$\sqrt{n}\left(\widehat{\beta} - \beta\right) \xrightarrow{D} N(0, \Sigma_1), \tag{2.6}$$

where $\Sigma_1 = A^{-1}D_1A^{-1}$.

To apply *Theorem 2.3* to construct the confidence region of β, we give the estimator of Σ_1, say $\widehat{\Sigma}_1 = \widehat{A}^{-1}\widehat{D}_1\widehat{A}^{-1}$, where $\widehat{A}$ and $\widehat{D}_1$ are defined by $\widehat{A} = (1/n)\sum_{i=1}^{n} X_i X_i^T$, $\widehat{D}_1 = (1/n)\sum_{i=1}^{n}(1/p(X, \widehat{\theta}))X_i X_i^T (Y_i - X_i'\widehat{\beta}_r)^2$.

It is easily proved that $\widehat{\Sigma}_1$ is a consistent estimator of Σ_1. Thus, by *Theorem 2.3*, we have

$$\widehat{\Sigma}_1^{-1/2}\sqrt{n}\left(\widehat{\beta} - \beta\right) \xrightarrow{D} N(0, I_p), \tag{2.7}$$

where I_p is an identity matrix of order p. Using (10.2d) in Arnold [6], we can obtain

$$\left(\widehat{\beta} - \beta\right)^T n\widehat{\Sigma}_1^{-1}\left(\widehat{\beta} - \beta\right) \xrightarrow{D} \chi_p^2. \tag{2.8}$$

Therefore, the confidence region of β can be constructed by using (2.8).

2.2. Adjusted Empirical Likelihood When $p_0(x) \neq p(x, \theta)$

In this subsection, we also construct the empirical likelihood statistic and discuss some asymptotic results when the selection probability is specified incorrectly, that is, $p_0(x) \neq p(x, \theta)$ for any θ.

Since the design variable X_i is observable for each subject, the quasi-maximum likelihood estimator $\tilde{\theta}$, other than the maximum likelihood estimator, can be obtained by maximizing the likelihood function

$$L(\theta) = \prod_{i=1}^{n} p(X_i, \theta)^{\delta_i} (1 - p(X_i, \theta))^{1-\delta_i} \tag{2.9}$$

under some regularity assumptions.

The following regularity assumptions are sufficiently strong to ensure both consistency and asymptotic normality of the quasi-maximum likelihood estimator. Let $u = (x, \delta)$, $\Omega = (R^p, \{0, 1\})$, $g(u) = p_0(x)^{\delta}(1 - p_0(x))^{1-\delta}$, $f(u, \theta) = p(x, \theta)^{\delta}(1 - p(x, \theta))^{1-\delta}$. It is natural that $g(u)$ is the true density function of u, and $f(u, \theta)$ does not contain the true structure $g(u)$. The Kullback-Leibler Information Criterion (KLIC) can be defined by $I(g : f, \theta) = E(\log[g(U)/f(U, \theta)])$, here, and in what follows, expectations are taken with respect to the true distribution $g(u)$. When expectations of the partial derivatives exist, we define the matrices $A(\theta) = E(\partial^2 \log f / \partial\theta_i\theta_j)$, $B(\theta) = E((\partial \log f / \partial\theta_i)(\partial \log f / \partial\theta_j))$.

(C7) $g(u)$ is measurable on Ω.

(C8) $f(u, \theta)$ are measurable in u for every θ in Θ, a compact subset of a R^p, and continuous u for every θ in Θ.

(C9) (a) $E[\log(g(U))]$ exists and $|\log f(u, \theta)| \leq m(u)$ for all θ in Θ, where m is integrable with respect to g; (b) $I(g : f, \theta)$ has a unique minimum at θ^* in Θ.

(C10) $E((\partial^2 \log f)/(\partial\theta_i))$, $i = 1, \ldots, p$, are measurable functions of u for each θ in Θ and continuously second-order differentiable functions of θ for each u in Ω.

(C11) $|\partial^2 \log f / \partial\theta_i\theta_j|$ and $|(\partial \log f / \partial\theta_i)(\partial \log f / \partial\theta_j)|$, $i, j = 1, \ldots, p$ are dominated by functions integrable with respect to g for all u in Ω and θ in Θ.

(C12) (a) θ^* is interior to Θ; (b) $B(\theta^*)$ is nonsingular; (c) θ^* is a regular point of $A(\theta^*)$.

Then, we use $\tilde{Y}_i = (\delta_i/p(X_i, \tilde{\theta}))Y_i + (1 - (\delta_i/p(X_i, \tilde{\theta})))X_i'\hat{\beta}_r$, $i = 1, 2, \ldots, n$ as "complete" data set for Y to construct the auxiliary random vectors

$$Z_{\text{in}}(\beta) = X_i(\tilde{Y}_i - X_i'\beta) = X_i \left\{ \frac{\delta_i}{p(X_i, \tilde{\theta})}(Y_i - X_i'\beta) + \left(1 - \frac{\delta_i}{p(X_i, \tilde{\theta})}\right)X_i'(\hat{\beta}_r - \beta) \right\}. \tag{2.10}$$

Thus, an empirical log-likelihood ratio is defined as

$$l_n(\beta) = -2 \max \left\{ \sum_{i=1}^{n} \log(n\omega_i) \mid \omega_i \geq 0, \sum_{i=1}^{n} \omega_i = 1, \sum_{i=1}^{n} \omega_i Z_{\text{in}}(\beta) = 0 \right\}. \tag{2.11}$$

And the maximum empirical likelihood estimator $\tilde{\beta}$ of β is to maximize $\{-l_n(\beta)\}$.

To ensure asymptotic results, the following assumptions are needed.

(C4′) $p(x, \theta)$ is uniformly continued in $U(\theta^*)$ for all x;

(C5′) A, C, E, F, G, D_1, D_2 is a positive definite matrix, where $A = E\{XX^T\}$, $C = E[XX^T(1 - p_0(X)/p(X, \theta^*))]$, $E = E[p_0(X)XX^T]$, $F = E[p_0(X)XX^T\varepsilon^2]$, $G = E[(p_0(X)/p(X, \theta^*))XX^T\varepsilon^2]$, $D_2 = E\{(p_0(X)/p^2(X, \theta^*))XX'\varepsilon^2\}$, and $D_3 = D_2 + CE^{-1}FE^{-1}C^T + 2CE^{-1}G$;

(C6′) $p_0(x)$ has bounded almost surely and $in f_x p_0(x) > 0$,

where condition (C4′) is common for some selection probability function. Condition (C5′)-(C6′) are necessary for asymptotic normality of the maximum empirical likelihood estimator.

Theorem 2.4. *Suppose that conditions (C7)–(C12) and (C4′)–(C6′) above hold. If β is the true parameter and $p(x, \theta)$ is specified incorrectly, then*

$$l_n(\beta) \xrightarrow{D} \sum_{i=1}^{p} \tilde{\omega}_i \chi^2_{1,i}, \tag{2.12}$$

where χ^2_1 means the chi-square variable with 1 degrees of freedom. The weights $\tilde{\omega}_i$, $i = 1, 2, \ldots, p$ are the eigenvalues of matrix $D_3^{-1}D_2$. $\xrightarrow{D}$ represents the convergence in distribution.

Let $r(\beta) = p/tr(D_3^{-1}D_2)$ be adjustment factor. Along the lines of [7], it is straightforward to show that $r(\beta)\Sigma_{i=1}^{p}\tilde{\omega}_i\chi^2_{1i} \xrightarrow{D} \chi^2_p$.

Let $S_{n1}(\beta) = \Sigma_{i=1}^{n}Z_{in}(\beta)Z_{in}^T(\beta)$, $S_{n2}(\beta) = (\Sigma_{i=1}^{n}Z_{in}(\beta))(\Sigma_{i=1}^{n}Z_{in}(\beta))^T$, $\hat{r}(\beta) = tr(\hat{S}_{n2}^{-1}(\beta)S_{n2}(\beta))/tr(\hat{S}_{n1}^{-1}(\beta)S_{n2}(\beta))$. We define an adjusted log-likelihood ratio by

$$l_{n,\text{ad}}(\beta) = \hat{r}(\beta)l_n(\beta). \tag{2.13}$$

Corollary 2.5. *Under the conditions of Theorem 2.4, one has*

$$l_{n,ad}(\beta) \xrightarrow{D} \chi^2_p. \tag{2.14}$$

Let the $\chi^2_p(1 - \alpha)$ be the $1 - \alpha$ quartile of the χ^2_p for $0 < \alpha < 1$. Using Corollary 2.5, we obtain an approximate $1 - \alpha$ confidence region for β, defined by

$$\left\{ \check{\beta} \mid l_{n,ad}\left(\check{\beta}\right) \leq \chi^2_p(1 - \alpha) \right\}. \tag{2.15}$$

Corollary 2.5 can also be used to test the hypothesis $H_0 : \beta = \beta_0$. One could reject H_0 if $l_{n,ad}(\check{\beta}) > \chi^2_p(1 - \alpha)$.

Note that $r(\beta) \to 1$, when $p(x, \theta)$ is close to correct one. Actually, the adjustment factor $r(\beta)$ reflects information loss due to the misspecification of $p(x, \theta)$.

Theorem 2.6. *Under the conditions of Theorem 2.4, one has*

$$\sqrt{n}\left(\tilde{\beta} - \beta\right) \xrightarrow{D} N(0, \Sigma_2), \tag{2.16}$$

where $\Sigma_2 = A^{-1} D_3 A^{-1}$.

To apply Theorem 2.6 to construct the confidence region of β, we give the estimator of Σ_2, say $\hat{\Sigma}_2 = \hat{A}^{-1} \hat{D}_3 \hat{A}^{-1}$, where $\hat{A}$ and $\hat{D}_3$ are defined by $\hat{A} = (1/n) \sum_{i=1}^{n} X_i X_i^T$, $\hat{C} = (1/n) \sum_{i=1}^{n} X_i X_i^T (1 - \delta_i / p(X_i, \theta^*))$, $\hat{E} = (1/n) \sum_{i=1}^{n} \delta_i X_i X_i^T$, $\hat{F} = (1/n) \sum_{i=1}^{n} \delta_i X_i X_i^T (Y_i - X_i' \hat{\beta}_r)^2$, $\hat{G} = (1/n) \sum_{i=1}^{n} (\delta_i / p(x, \theta^*)) X_i X_i^T (Y_i - X_i' \hat{\beta}_r)^2$, $\hat{D}_2 = (1/n) \sum_{i=1}^{n} (\delta_i / p^2(x, \theta^*)) X_i X_i^T (Y_i - X_i' \hat{\beta}_r)^2)$, $\hat{D}_3 = \hat{D}_2 + \hat{C} \hat{E}^{-1} \hat{F} \hat{E}^{-1} \hat{C}^T + 2 \hat{C} \hat{E}^{-1} \hat{G}$.

It is easily proved that $\hat{\Sigma}_2$ is a consistent estimator of Σ_2. Thus, by Theorem 2.6, we have

$$\hat{\Sigma}_2^{-1/2} \sqrt{n}\left(\tilde{\beta} - \beta\right) \xrightarrow{D} N(0, I_p), \tag{2.17}$$

where I_p is an identity matrix of order p. Using (10.2d) in Arnold [6], we can obtain

$$\left(\tilde{\beta} - \beta\right)^T n \hat{\Sigma}_2^{-1} \left(\tilde{\beta} - \beta\right) \xrightarrow{D} \chi_p^2. \tag{2.18}$$

Therefore, the confidence region of β can be constructed by using (2.18).

Remark 2.7. The estimator proposed by Robins in this situation is to solve

$$\sum_{i=1}^{n} \frac{\delta_i}{p\left(X_i, \hat{\theta}\right)} X_i(Y_i - X_i' \beta) + \left(1 - \frac{\delta_i}{p\left(X_i, \hat{\theta}\right)}\right) X_i X_i'\left(\hat{\beta}_r - \beta\right) = 0. \tag{2.19}$$

Noted that $E(X(Y - X'\beta) \mid X) = XX'(\beta - \beta) = XX'[(\hat{\beta}_r - \beta) + (\beta - \hat{\beta}_r)] = XX'(\hat{\beta}_r - \beta) + o_p(1/\sqrt{n})$ that is, in that case the underlying regression function can asymptotically correctly specified. So whether missing data mechanism is specified correctly or not, the estimator is always consistent and the estimator can achieve asymptotic full efficiency when specified correctly.

3. Simulation

Due to the curse of nonparametric estimation, Xue's method may be hard to realize. Here we conducted an extensive simulation study to compare the performances of the weighted-corrected empirical likelihood(WCEL) proposed in this paper and Wang's method (AEL) proposed in Wang and Rao under the covariates of four dimensions.

We considered the linear model (1.1) with $d = 4$ and $\beta = (0.8, 1.5, 1, 2)$, where X was generated from a four-dimensions standard normal distribution, and ε was generated from the normal distribution with mean zero and variance 0.04.

Table 1: Empirical coverage probabilities of the confidence regions for β under different selection probability functions $p(x)$ and sample sizes n when nominal level is 0.95.

	n	Case I			Case II		
		$p_0(x,\theta_1)$	$p_0(x,\theta_2)$	$p_0(x,\theta_3)$	$p_1(x)$	$p_2(x)$	$p_3(x)$
	100	0.554	0.61	0.571	0.861	0.772	0.693
WCEL	200	0.594	0.729	0.611	0.971	0.916	0.818
	500	0.699	0.914	0.698	0.995	0.987	0.967
	100	0.486	0.489	0.516	0.667	0.591	0.521
AEL	200	0.549	0.536	0.564	0.805	0.587	0.587
	500	0.655	0.653	0.660	0.932	0.863	0.759

In the first case, the real selection probability function $p_0(x,\theta)$ was taken to be $\exp(x'\theta)/(1 + \exp(x'\theta))$. We considered θ equaled to the three following values $\theta_1 = (-0.5, -0.5, -0.5, -0.5)$, $\theta_2 = (0, 0, 0, 0)$, $\theta_3 = (0.5, 0.5, 0.5, 0.5)$, respectively.

In the second case, the real selection probability function $p_0(x)$ was taken to be the following three cases:

Case 1. $p_1(x) = 0.8 + 0.2(|x_1| + |x_2| + |x_3| + |x_4|)$ if $|x_1| + |x_2| + |x_3| + |x_4| \leq 1$, and 0.9 elsewhere.

Case 2. $p_2(x) = 0.9 - 0.1(|x_1| + |x_2| + |x_3| + |x_4|)|$ if $|x_1| + |x_2| + |x_3| + |x_4| \leq 4$, and 0.9 elsewhere.

Case 3. $p_3(x) = 0.6$ for all x.

We generated 5000 Monte Carlo random samples of size n = 100, 200, and 500 based on the above six selection probability functions $p(x)$. When the working model was $p(x,\theta) = \exp(x'\theta)/(1 + \exp(x'\theta))$, the empirical coverage probabilities for β, with a nominal level 0.95, were computed based on the above two methods with 5000 simulation runs. The results are reported in Table 1.

From Table 1, we can obtain the following results. Firstly, under both cases, WCEL performs better than AEL because its confidence regions have uniformly higher coverage probabilities. Secondly, all the empirical coverage probabilities increase as n increases for every fixed missing rate. Observably, the missing rate also affects coverage probability. Generally, the coverage probability decreases as the missing rate increases for every fixed sample size. However, under Case I, the values do hardly change by a large amount for both methods because of the exponential selection probability function.

4. Conclusion

In this paper, a parametric selection probability function is introduced to avoid the dimension difficulty, and a bias-corrected technique leads to an empirical likelihood (EL) statistic with asymptotically chi-square distribution when the selection probability specified correctly and with asymptotically weighted chi-square distribution when specified incorrectly. Also, our estimator is always consistent and will achieve asymptotic full efficiency when selection probability function is specified correctly.

Appendix

A. Proofs

Lemma A.1. *Suppose that (C1)–(C4) hold. Then $1/p(X, \widehat{\theta}) = (1/p(X, \theta_0))(1 + o_p(1))$.*

Proof of Lemma A.1. Under conditions (C1)–(C3), it can be shown that $\widehat{\theta}$ is $\sqrt{n}$ consistent:

$$\widehat{\theta} - \theta_0 = O_p\left(n^{-1/2}\right) = o_p(1). \tag{A.1}$$

Together with condition (C4), we have

$$p\left(X, \widehat{\theta}\right) = p(X, \theta_0) + o_p(1). \tag{A.2}$$

By Taylor expanding, it is easily shown that

$$\frac{1}{p\left(X, \widehat{\theta}\right)} = \frac{1}{p(X, \theta_0)}\left(1 + o_p(1)\right). \tag{A.3}$$

□

Lemma A.2. *Suppose that (C7)–(C12) hold. Then $1/p(X, \widetilde{\theta}) = (1/p(X, \theta^*))(1 + o_p(1))$.*

Proof of Lemma A.2. From Theorem 3 of White [8], under conditions (C7)–(C12), it can be shown that $\widetilde{\theta}$ is $\sqrt{n}$ consistent:

$$\widetilde{\theta} - \theta^* = O_p\left(n^{-1/2}\right) = o_p(1). \tag{A.4}$$

Similar to Lemma A.1, it is easy to show that $1/p(X, \widetilde{\theta}) = (1/p(X, \theta^*))(1 + o_p(1))$. □

Lemma A.3. *Suppose that (C7)–(C12) and (C4′)–(C6′) hold. If $p(X, \theta)$ is specified wrongly, then*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_{in}(\beta) \longrightarrow N(0, D_3), \tag{A.5}$$

$$\frac{1}{n} \sum_{i=1}^{n} Z_{in}(\beta) Z_{in}^{T}(\beta) \longrightarrow D_2,$$

$$\max |Z_{in}(\beta)| = o\left(n^{-1/2}\right). \tag{A.6}$$

Proof of Lemma A.3. We prove (A.5) only; (A.6) can be proved similarly.

By direct calculation, we have

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_{\text{in}}(\beta) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{\delta_i}{p\left(X_i,\widetilde{\theta}\right)} X_i(Y_i - X_i'\beta) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(1 - \frac{\delta_i}{p\left(X_i,\widetilde{\theta}\right)}\right) X_i X_i'\left(\widehat{\beta}_r - \beta\right) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{\delta_i}{p(X_i,\theta^*)} X_i \varepsilon_i (1 + o_p(1)) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\left(1 - \frac{\delta_i}{p(X_i,\theta^*)}\right) X_i X_i'\left(\frac{1}{n}\sum_{i=1}^{n}\delta_i X_i X_i'\right)^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_i X_i \varepsilon_i \\
&\quad + \frac{1}{n}\sum_{i=1}^{n} X_i \frac{\delta_i}{p(X_i,\theta^*)} X_i'\left(\frac{1}{n}\sum_{i=1}^{n}\delta_i X_i X_i'\right)^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_i X_i \varepsilon_i o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{\delta_i}{p(X_i,\theta^*)} X_i \varepsilon_i \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\left(1 - \frac{\delta_i}{p(X_i,\theta^*)}\right) X_i X_i'\left(\frac{1}{n}\sum_{i=1}^{n}\delta_i X_i X_i'\right)^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_i X_i \varepsilon_i + o_p(1).
\end{aligned}
\tag{A.7}
$$

It is easily to shown that

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{\delta_i}{p(X_i,\theta^*)} X_i \varepsilon_i &\longrightarrow N(0, D_2), \\
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_i X_i \varepsilon_i &\longrightarrow N(0, F), \\
\frac{1}{n}\sum_{i=1}^{n}\left(1 - \frac{\delta_i}{p(X_i,\theta^*)}\right) X_i X_i' &= C + o_p(1), \\
\frac{1}{n}\sum_{i=1}^{n}\delta_i X_i X_i' &= E + o_p(1).
\end{aligned}
\tag{A.8}
$$

Note that

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left(1 - \frac{\delta_i}{p(X_i,\theta^*)}\right) X_i X_i'\left(\frac{1}{n}\sum_{i=1}^{n}\delta_i X_i X_i'\right)^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_i X_i \varepsilon_i \\
&\quad = \left(\frac{1}{n}\sum_{i=1}^{n}\left(1 - \frac{\delta_i}{p(X_i,\theta^*)}\right) X_i X_i' - C\right)\left(\frac{1}{n}\sum_{i=1}^{n}\delta_i X_i X_i'\right)^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_i X_i \varepsilon_i \\
&\quad\quad + C\left(\left(\frac{1}{n}\sum_{i=1}^{n}\delta_i X_i X_i'\right)^{-1} - E^{-1}\right) \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_i X_i \varepsilon_i + CE^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_i X_i \varepsilon_i
\end{aligned}
$$

$$= CE^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_i X_i \varepsilon_i + o_p(1),$$

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_i X_i \varepsilon_i \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i}{p(X_i, \theta^*)} X_i \varepsilon_i \right)' = G + o_p(1),$$

(A.9)

so we have

$$\frac{1}{n} \sum_{i=1}^{n} \left(1 - \frac{\delta_i}{p(X_i, \theta^*)} \right) X_i X_i' \left(\frac{1}{n} \sum_{i=1}^{n} \delta_i X_i X_i' \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_i X_i \varepsilon_i \longrightarrow N \left(0, CE^{-1} FE^{-1} C^T \right),$$

$$\frac{1}{n} \sum_{i=1}^{n} \left(1 - \frac{\delta_i}{p(X_i, \theta^*)} \right) X_i X_i' \left(\frac{1}{n} \sum_{i=1}^{n} \delta_i X_i X_i' \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_i X_i \varepsilon_i \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i}{p(X_i, \theta^*)} X_i \varepsilon_i \right)'$$

(A.10)

$$= CE^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_i X_i \varepsilon_i \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i}{p(X_i, \theta^*)} X_i \varepsilon_i \right)' + o_p(1)$$

$$= CE^{-1} G + o_p(1).$$

Therefore, by using (A.8)–(A.10), the proof of (A.5) is completed. □

We just proof Theorems 2.4 and 2.6, the proofs of Theorems 2.1 and 2.3 are similar, only to replace the $\tilde{\theta}$, θ^*, D_3, $Z_{in}(\beta)$, C by $\hat{\theta}$, θ_0, D_1, $Z_{in}^*(\beta)$, 0 separately.

Proof of Theorem 2.4. By the Lagrange multiplier method, $Z_{in}(\beta)$ can be represented as

$$l_n(\theta) = 2 \sum_{i=1}^{n} \log \left(1 + \lambda^T Z_{in}(\beta) \right),$$

(A.11)

where $\lambda = \lambda(\beta)$ is a $d \times a$ vector given as the solution to

$$\sum_{i=1}^{n} \frac{Z_{in}(\beta)}{1 + \lambda^T Z_{in}(\beta)} = 0,$$

(A.12)

by Lemma A.2, and using the same arguments as that of the proof (A.4) of in [9], we can show that

$$\|\lambda\| = O_p \left(n^{-1/2} \right).$$

(A.13)

Applying the Taylor expansion to (A.11) and (A.13), we get that

$$l_n(\theta) = 2\sum_{i=1}^{n}\left[\lambda^T Z_{in}(\beta) - \frac{\lambda^T Z_{in}(\beta)^2}{2}\right] + o(1). \tag{A.14}$$

By (A.12), it follows that

$$0 = \sum_{i=1}^{n}\frac{Z_{in}(\beta)}{1 + \lambda^T Z_{in}(\beta)} = \sum_{i=1}^{n}Z_{in}(\beta) - \sum_{i=1}^{n}Z_{in}(\beta)\lambda^T Z_{in}(\beta) + o_p\left(n^{-1/2}\right). \tag{A.15}$$

This together with Lemma A.3 and (A.13) proves that

$$\sum_{i=1}^{n}\lambda^T Z_{in}(\beta)^2 = \sum_{i=1}^{n}\lambda^T Z_{in}(\beta) + o(1),$$

$$\lambda = \sum_{i=1}^{n}\left(Z_{in}(\beta)Z_{in}^T(\beta)\right)^{-1}\sum_{i=1}^{n}Z_{in}(\beta) + o_p\left(n^{-1/2}\right). \tag{A.16}$$

Therefore, from (A.14) we have

$$l_n(\beta) = \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Z_{in}(\beta)\right)^T\left(\frac{1}{n}\sum_{i=1}^{n}Z_{in}(\beta)Z_{in}^T(\beta)\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Z_{in}(\beta)\right) + o(1). \tag{A.17}$$

This together with Lemma A.3 completes Theorem 2.4.   □

Proof of Theorem 2.6. From Theorem 1 of Qin [10] and (A.5), we obtain the result of Theorem 2.6 directly.   □

Acknowledgment

References

[1] A. B. Owen, *Empirical Likelihood*, Chapman and Hall, New York, NY, USA, 2001.

[2] Q. Wang and J. N. K. Rao, "Empirical likelihood-based inference in linear models with missing data," *Scandinavian Journal of Statistics*, vol. 29, no. 3, pp. 563–576, 2002.

[3] L. Xue, "Empirical likelihood for linear models with missing responses," *Journal of Multivariate Analysis*, vol. 100, no. 7, pp. 1353–1366, 2009.

[4] Y. S. Qin, L. Li, and A. Z. Lei, "Empirical likelihood for linear regression models with missing responses," *Statistics & Probability Letters*, vol. 79, no. 11, pp. 1391–1396, 2009.

[5] J. M. Robins, A. Rotnitzky, and L. P. Zhao, "Estimation of regression coefficients when some regressors are not always observed," *Journal of the American Statistical Association*, vol. 89, no. 427, pp. 846–866, 1994.

[6] S. F. Arnold, *The Theory of Linear Models and Multivariate Analysis*, John Wiley & Sons, New York, NY, USA, 1981.

[7] J. N. K. Rao and A. J. Scott, "The analysis of categorical data from complex sample surveys: chi-squared tests for goodness of fit and independence in two-way tables," *Journal of the American Statistical Association*, vol. 76, no. 374, pp. 221–230, 1981.

[8] H. White, "Maximum likelihood estimation of misspecified models," *Econometrica*, vol. 50, no. 1, pp. 1–25, 1982.

[9] A. Owen, "Empirical likelihood for linear models," *The Annals of Statistics*, vol. 19, no. 4, pp. 1725–1747, 1991.

[10] J. Qin and J. Lawless, "Empirical likelihood and general estimating equations," *The Annals of Statistics*, vol. 22, no. 1, pp. 300–325, 1994.