Research Article

Mean Empirical Likelihood Inference for Response Mean with Data Missing at Random

Hanji He ¹ and Guangming Deng ^{1,2}

¹College of Science, Guilin University of Technology, Guilin 541004, China

²Applied Statistics Institute, Guilin University of Technology, Guilin 541004, China

Correspondence should be addressed to Guangming Deng; dgm@glut.edu.cn

Received 14 September 2020; Accepted 27 November 2020; Published 15 December 2020

Academic Editor: Stefan Balint

Copyright © 2020 Hanji He and Guangming Deng. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We extend the mean empirical likelihood inference for response mean with data missing at random. The empirical likelihood ratio confidence regions are poor when the response is missing at random, especially when the covariate is high-dimensional and the sample size is small. Hence, we develop three bias-corrected mean empirical likelihood approaches to obtain efficient inference for response mean. As to three bias-corrected estimating equations, we get a new set by producing a pairwise-mean dataset. The method can increase the size of the sample for estimation and reduce the impact of the dimensional curse. Consistency and asymptotic normality of the maximum mean empirical likelihood estimators are established. The finite sample performance of the proposed estimators is presented through simulation, and an application to the Boston Housing dataset is shown.

1. Introduction

The missing data problems exist widely in the social sciences, political sciences, medical research, and many other fields. There are different missing patterns including single variable nonresponse, multivariate nonresponse, monotone nonresponse, and general nonresponse, and there are three types of missing mechanisms: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). The MAR is defined as missing that only depends on the observed. Let Y be the value of a response that is subject to nonresponse and let X be a d -dimensional vector of auxiliary variables that are observed fully. Y_o means the dataset of Y which is observed, and Y_m means the dataset of Y which is missing or nonresponse. We interested in the mean of Y ; i.e., $\mu = E(y)$. Let M be the indicator for Y , where $M = 1$ if Y is observed and $M = 0$ otherwise. In the MAR mechanism,

$$L(A|Y, X) = L(A|Y_o, X), \quad (1)$$

means that missing data of Y is only related to (Y_o, X) and has nothing to do with Y_m .

Empirical likelihood [1, 2] which is widely used for nonparametric and semiparametric statistical inferences is a competitive and powerful method for constructing confidence intervals (CIs). The empirical log-likelihood ratio (ELR) usually has an asymptotic chi-squared distribution, and the CI based on EL has many excellent properties, as proposed by Owen [3]. Due to these excellent properties, Wang and Rao [4], Liang et al. [5], and Stute et al. [6] extended the EL to missing data by imputing the missing data by a kernel regression function of observed data. For example, the bias-corrected EL method was constructed via the inverse propensity weighting (IPW), mean imputation (MI), and augmented inverse propensity weighting (AIPW). Guan et al. [7] proposed that an efficient estimator of the response means is constructed using the estimated response probabilities, which is a new method based on semiparametric maximum likelihood inference. Xie and Zhang [8] construct empirical likelihood confidence regions for the regression parameters without and with constraints by unconstrained and constrained empirical likelihood ratio statistics. Wang and Deng [9] proposed dimension-reduced empirical likelihood that was a two-stage estimation

procedure and applied sufficient dimension reduction (SDR) technique in the kernel estimation of the propensity as well as the conditional mean response function. The method can avoid the well-known curse of dimensionality in the multivariate situation. But, it works only with a large sample. A lot of studies devote themselves to improve the accuracy of empirical likelihood ratio confidence region inference when there are small sample sizes and multidimensional situations. Hall and Scala [10], Diccio et al. [11], and Tsao [12] found that empirical likelihood ratio confidence intervals could have poor accuracy, especially in the small sample and multidimensional situations. For confidence estimation, Diccio et al. [11], Chen et al. [13], and Abraham and Wu [14] proposed the Bartlett correction empirical likelihood (BEL), the adjusted empirical likelihood, and an extended empirical likelihood (EEL), respectively. However, these methods require the calculation of the Bartlett correction constant, which is difficult to compute. So, Liang et al. [15] proposed the new method named mean empirical likelihood, which constructs an empirical likelihood function based on a set of pseudodata and it is easy to compute. Therefore, we need to develop a new empirical likelihood inference method for response mean with data missing at random in case of a small sample and multidimensional data.

Let V be the pairwise-mean dataset. We will introduce a new empirical likelihood method named mean empirical likelihood. It constructs an empirical likelihood function based on a set of pseudodata and it is easy to complete compared to the mean empirical likelihood (MEL). The large sample properties of MEL are presented by Liang et al. [15]. The simulation studies indicate that the confidence regions by MEL are much more accurate than those by other empirical likelihood methods. Bigger coverage probability and smaller average interval length are other significant advantages of MEL. Nonetheless, there are two main problems in inference for response mean with data missing at random. First, the size of the sample is too small so that asymptotic normality is a little hard. The other is the high dimension of the covariate, and in these cases, the EL method is ineffective. MEL can solve the above problems by increasing the size of the sample for estimation.

In this paper, we propose a new empirical likelihood to make efficient inference on response mean, especially for confidence region of response mean, by applying the mean empirical likelihood inference for small sample sizes and multidimensional variable situations. In specific, we first apply the mean empirical likelihood (MEL) method (see [15]) for response mean with data missing at random. We construct three bias-corrected nonparametric MEL functions by pairwise-mean dataset based on the IPW, MI, and AIPW approaches. We showed that the resulting three MEL methods yield asymptotically equivalent estimators that achieve the desirable asymptotic unbiasedness and asymptotic normality. After a factor adjusted to the MELR function based on the IPW or MI, the proposed three MELR functions are shown to be asymptotically standard chi-squared with $(d-1)$ degree of freedom, where d is the rank of the

covariance matrix of the pair-mean dataset. Simulation results show that the proposed method not only has higher accurate coverage probabilities, smaller average lengths, and standard deviation but also has efficient point estimators when the sample size is only 40.

This paper is organized as follows. Section 2 presents three types of nonparametric MELRs. Then, we introduce our main idea and establish a number of asymptotic properties. The simulation studies are given in Section 3. Section 4 provides a real data analysis and the paper concludes with a discussion in Section 5.

2. Mean Empirical Likelihood Inference for Response Mean with Data Missing at Random

2.1. Some Methods for Missing Response Problem. Let (X_i, Y_i, M_i) , $i = 1, \dots, n$, be independent and identically distributed realizations from (X, Y, M) , where X_i is a fully observed multivariate covariate, Y_i is a univariate response having missing values, and M_i is a binary response indicator that equals 1 if and only if Y_i is observed.

In model (1), we make MAR assumptions about the full data, and the missing pattern is generally nonresponse. Furthermore, To adjust the weight, we introduce the probability function $\pi(X)$, which is $\pi(x) = \Pr(M = 1|X, Y) = \Pr(M = 1|X)$, and conditional mean response function, such that $m(X) = E(Y|X, M = 1) = (E(MY|X)/E(M|X))$. The unspecified functions $\pi(X)$ and $m(X)$ can be estimated by the kernel regression estimators as follows:

$$\begin{aligned}\hat{\pi}(X) &= \frac{\sum_{j=1}^n M_j K_h(X - X_j)}{\sum_{j=1}^n K_h(X - X_j)}, \\ \hat{m}(X) &= \frac{\sum_{j=1}^n M_j Y_j K_h(X - X_j)}{\sum_{j=1}^n \delta_j K_h(X - X_j)},\end{aligned}\tag{2}$$

where $K_h(\bullet) = h^{-1}K(\bullet/h)$, $K(\bullet)$ is a symmetric kernel function, and h is a bandwidth.

Next, we introduce three bias-corrected estimating equations using the nonparametric kernel estimators $\hat{\pi}(X)$ and $\hat{m}(X)$ in equation (2) for handling missing data as follows.

- (i) Nonparametric inverse propensity weighting (IPW) approach: this method assigns each observed Y_i with weight proportional to the inverse of the estimated propensity $\hat{\pi}(X_i)$ in (2), i.e.,

$$\hat{\varphi}_{1i}(\mu) = \frac{M_i Y_i}{\hat{\pi}(X_i)} - \mu, \quad i = 1, \dots, n.\tag{3}$$

- (ii) Nonparametric mean imputation (MI) approach: for each X_i with $\delta_i = 0$, we use an estimator $\hat{m}(X_i)$ in equation (2) to estimate $m(X_i)$, and a nonparametric

imputation estimating equation for μ is given by

$$\hat{\varphi}_{2i}(\mu) = M_i Y_i + (1 - M_i) \hat{m}(X_i) - \mu, \quad (4)$$

$$i = 1, \dots, n.$$

(iii) Nonparametric augmented inverse propensity weighting (AIPW) approach: we combine the nonparametric IPW and MI approaches leading to the nonparametric AIPW estimating equation for μ as follows:

$$\hat{\varphi}_{3i}(\mu) = \frac{M_i Y_i}{\hat{\pi}(X_i)} + \left\{ 1 - \frac{M_i}{\hat{\pi}(X_i)} \right\} \hat{m}(X_i) - \mu, \quad (5)$$

$$i = 1, \dots, n.$$

In general, we use the empirical likelihood method to estimate the response mean. Let p_{li} present the probability weight allocated to $\hat{\varphi}_{li}(\mu)$ for $l = 1, 2, 3$, $i = 1, \dots, n$. The profile empirical log-likelihood ratio function for μ based on $\hat{\varphi}_{li}(\mu)$ with data missing at random is defined as

$$\hat{R}_l(\mu) = -2 \sup \left\{ \sum_{i=1}^n \log(np_{li}) \mid p_{li} \geq 0, \sum_{i=1}^n p_{li} = 1, \sum_{i=1}^n p_{li} \hat{\varphi}_{li}(\mu) = 0 \right\}. \quad (6)$$

However, when the sample size is less than 100 or the data is high dimensional, its finite sample properties may not work well because of the low precision of χ^2 approximation. Hence, the efficiency of the EL method for the response mean with data missing at random is low. In particular, the estimated coverage probability and average length are far from the theoretical value. The mean empirical likelihood can deal with a small sample and multidimensional situations of response mean problem with data missing at random.

2.2. Mean Empirical Likelihood for Response Mean with Data Missing at Random. In this paper, we apply the MEL to estimate the response mean with data missing at random under small sample and high-dimensional variate situation.

We here follow notations in the previous section, and for simplicity, we denote $\phi_{li}(\mu) = \phi_{li}(Y_i, \mu) = g(X_i, \theta)$, $i = 1, \dots, n$, and further denote the pairwise-mean dataset as follows:

$$\nu = \left\{ \frac{\phi_{li}(\mu) + \phi_{lj}(\mu)}{2} : 1 \leq i \leq j \leq n \right\}, \quad (7)$$

which can also be written as $\nu = \{V_{l1}(\mu), V_{l2}(\mu), \dots, V_{lN}(\mu)\}$ with $N = n(n+1)/2$. Let p_{lk} donate the probability weight allocated to $\hat{\varphi}_{li}(\mu)$ for $l = 1, 2, 3, i = 1, \dots, n$. Based on this pairwise-mean dataset, the empirical log-likelihood ratio for μ is defined as

$$R_l^M(\mu) = -2 \sup \left\{ \sum_{k=1}^N \log(Np_{lk}) \mid \sum_{k=1}^N p_{lk} V_{lk}(\mu) = 0, \sum_{k=1}^N p_{lk} = 1, p_{lk} \geq 0, k = 1, \dots, N \right\}, \quad (8)$$

which is named mean empirical likelihood ratio. It follows that

$$R_l^M(\mu) = -2 \sum_{k=1}^N \log(Np_{lk}) = 2 \sum_{k=1}^N \log\{1 + \lambda_l' V_{lk}(\mu)\}, \quad (9)$$

where λ satisfying

$$\frac{1}{N} \sum_{k=1}^N \frac{V_{lk}(\mu)}{1 + \lambda_l' V_{lk}(\mu)} = 0, \quad (10)$$

$$p_{lk} = \frac{1}{N(1 + \lambda_l' V_{lk}(\mu))}.$$

Let θ_0 be the unknown true parameter value. Then, the mean empirical log-likelihood ratio is given by

$$L_l^M(\mu_0) = \frac{-2 \log R_l^M(\mu_0)}{(n+1)}. \quad (11)$$

Now, we have the following main theorem.

Theorem 1. Under the conditions listed in the Appendix, as $n \rightarrow \infty$, we have

$$\sqrt{n}(\hat{\mu}_l - \mu_0) \xrightarrow{L} N(0, \sigma^2), \quad (12)$$

where $l = 1, 2, 3$ denotes IPW, MI, and AIPW, respectively, and $\xrightarrow{L}$ denotes convergence in distribution.

Theorem 2. Under the conditions listed in the Appendix, as $n \rightarrow \infty$, we have

$$\begin{aligned} \hat{R}_{ipw}(\mu_0) &\xrightarrow{L} 2\vartheta_1 \chi_1^2, \\ \hat{R}_{mi}(\mu_0) &\xrightarrow{L} 2\vartheta_2 \chi_1^2, \\ \hat{R}_{aipw}(\mu_0) &\xrightarrow{L} 2\chi_1^2, \end{aligned} \quad (13)$$

where χ_1^2 is the χ^2 -distributed random variable with one degree of freedom, $\vartheta_1 = \sigma^2/\sigma_1^2$, $\vartheta_2 = \sigma^2/\sigma_2^2$.

Theorem 3. Under the conditions that $\text{Cov}(\phi_i(\mu_0)) = \Sigma$ exists and $\text{rank}(\Sigma) = d$, we have $L^M(\mu_0) \xrightarrow{L} \chi^2(d)$, in dist.

Proof. See Appendix. $\square$

Following Theorem 1, a confidence region for the parameter θ with asymptotic coverage probability $1 - \alpha$ can be defined as

$$I^M = \{\mu : L^M(\mu) \leq \chi_\alpha^2(d)\}. \quad (14)$$

3. Simulation Studies

We present two simulation studies in this section, which correspond to the finite sample performance of different MEL methods in Section 2. We will compare the MEL with OEL for response mean with data missing at random. For the one-dimensional covariate model, the estimating equation $g(X, \theta)$ is equal to the estimating equation $\phi_{li}(Y_i, \mu)$. For the multivariate covariate model, we compare the performance for different $\pi(X_i)$, correlation structures, and the size of the sample.

3.1. Simulation 1: Single Covariate Model. The results and discussion may be presented separately, or in one combined section, and may optionally be divided into headed sub-sections. Suppose that $g(X, \theta) = \phi_l(Y, \mu) = Y - \mu$ is the estimating equation for μ . We aim to compare the confidence intervals derived from MEL and OEL, for a given sample size n . We consider different scenarios by generating observations $X_1, X_2, \dots, X_n$ from $N(0, 1)$, respectively.

There are four indicators for evaluating method performance. The bias represents the difference between the estimated result and the true value, which measures the accuracy of point estimators. The coverage probability represents the probability that the point estimator falls within the confidence interval of the true value, which measures the stability of point estimators. The average length represents the average length of the confidence interval of the estimated value, and the variance represents the degree of dispersion of the estimated value, which measures the accuracy of interval estimators.

Based on 5000 replicates, the bias, coverage proportions, average length, and standard deviation were calculated. The simulation results are presented in Table 1.

Table 1 shows the confidence region characteristics of different empirical likelihood methods when the X_i is the one-dimensional normal distribution with mean 1 and variate 1, denoted as $X_i \sim N(1, 1)$. $Y_i = X_i + \varepsilon_i$ with $\varepsilon_i \sim N(0, 1)$ for $i = 1, 2, \dots, n$. Then, we generate δ_i from the Bernoulli distribution with probability function $\pi(X_i) = 1/(1 + \exp(-0.6 - 0.2X_i))$.

From Tables 1 and 2, we conclude the following:

- (1) Comparison for different sample sizes: when the sample sizes increase, all coverage probability estimated by OEL or MEL increases for small sample size n and all coverage probability estimated by OEL or MEL decreases. However, the bias estimated by AIPW and MI method reduced significantly, and the bias estimated by the complete-case and IPW method changed a little
- (2) Comparison of different estimation methods: MEL is much better than OEL. All coverage probabilities estimated by MEL are closer to the nominal levels than those of the OEL method. All standard deviations by MEL are much smaller than those of the OEL method. All average lengths by MEL are much smaller than those of the OEL method
- (3) Comparison for different estimating equations: the point estimator based on MI is superior to other estimating equations. In particular, the coverage probabilities estimated by the MI method perform very well in all simulations. In a word, the sequence about the accuracy of the estimating equation is that the MI method is superior to the AIPW method and the AIPW method is superior to IPW
- (4) Comparison for different error situations: the average length estimated by OEL and MEL with $\varepsilon_i \sim N(0, 1 + 0.5X_i^2)$ is much longer and more stable than the method with $\varepsilon_i \sim N(0, 1)$. But the standard deviation estimated by OEL and MEL with

$\varepsilon_i \sim N(0, 1 + 0.5X_i^2)$ is much bigger than the method with $\varepsilon_i \sim N(0, 1)$. Hence, when the errors in the method are heteroscedastic with $\varepsilon_i \sim N(0, 1 + 0.5X_i^2)$, MEL method for response mean also plays well

In a word, the performances of IPW, MI, and AIPW based on MEL precede those methods based on OEL, and AIPW based on MEL achieves the most accurate results.

3.2. Simulation 2: Multivariate Covariate Model. In simulation 2, X_i 's are 6-dimensional normal distribution with mean 0 and covariance matrix Γ , denoted by $X_i \sim N(0, \Gamma)$, where Γ is a 6×6 symmetric matrix with $\Gamma_{jj} = 1$ and $\Gamma_{jj'} = \rho^{|j' - j|}$ for $1 \leq j < j' \leq 6$, $Y_i = X_{i1} + X_{i2} + \dots + X_{i6} + \varepsilon_i$ with $\varepsilon_i \sim N(0, \sigma^2(X_i))$ for $i = 1, \dots, n$. Two kinds of ε_i 's are considered:

- (i) The first generates homoscedastic errors where ε_i and X_i are independent and $\sigma^2(X_i) = 1$
- (ii) The second errors are heteroscedastic with $\sigma^2(X_i) = 1 + 0.5X_{i1}^2$, $i = 1, 2, \dots, n$

We generate δ_i from the Bernoulli distribution with probability $\pi(X_i)$ and consider three choices of $\pi(X_i)$:

$$M1. \pi(X_i) = \frac{1}{[1 + \exp\{-0.6 - 0.2(X_{i1} + X_{i2} + X_{i3})\}]},$$

$$M2. \pi(X_i) = \frac{1}{[1 + \exp\{-0.2 - 0.4 \exp(X_{i1} + X_{i2} + X_{i3})\}]},$$

$$M3. \pi(X_i) = 1 - \exp[-\exp\{0.1 + 0.2(X_{i1} + X_{i2} + X_{i3})\}]. \quad (15)$$

The coefficients in the propensity models are chosen so that the unconditional rates of missing data are between 20% and 40%. When Y_i are linear and $\pi(X_i)$ is logistic linear under M1–M3, so S_Y and S_δ under M1–M3 are one-dimensional.

In addition, we also conduct two simulations with four correlation structures:

- (i) Independent covariates, where the correlation coefficient is 0
- (ii) Slightly correlated covariates, where the correlation coefficient is 0.2
- (iii) Moderately correlated covariates, where the correlation coefficient is 0.5
- (iv) Strongly correlated covariates, where the correlation coefficient is 0.8

All results are based on 1000 replications and the sample sizes are $n = 20$ and $n = 40$. Tables 3 and 4 show the biases and standard deviations (SDs) of the point estimators, the average lengths (ALs), and the coverage probabilities (CPs) of confidence intervals (CIs) at the nominal level of 95%. The CIs based on the mean empirical likelihood ratio $L^M(\mu)$ are obtained by I^M in equation (15). The CIs based on $\hat{\mu} = \bar{Y}$ and μ_{cc} are obtained by the normal approximation

TABLE 1: Performance of the point estimators and 95% CIs for response mean in simulation 1, $\varepsilon_i \sim N(0, 1)$.

Model		OEL					MEL				
		$\hat{\mu}_{ipw}$	$\hat{\mu}_{mi}$	$\hat{\mu}_{aipw}$	$\bar{Y}$	$\hat{\mu}_{cc}$	$\hat{\mu}_{ipw}$	$\hat{\mu}_{mi}$	$\hat{\mu}_{aipw}$	$\bar{Y}$	$\hat{\mu}_{cc}$
$n = 20$	Bias	-0.0106	0.0152	0.0098	0.0000	0.0586	-0.0106	0.0152	0.0098	0.0000	0.0586
	CP	0.9115	0.9130	0.9155	0.9135	0.9085	0.9290	0.9470	0.9375	0.9135	0.9085
	AL	1.2777	1.3051	1.2999	1.1771	1.4263	1.1764	1.2607	1.2212	1.1771	1.4263
	SD	1.7154	1.1895	1.4298	0.5447	0.5978	1.2415	0.8609	1.0348	0.5447	0.5978
$n = 30$	Bias	-0.0098	0.0088	0.0030	-0.0057	0.0549	-0.0098	0.0088	0.0030	-0.0057	0.0549
	CP	0.9285	0.9285	0.9295	0.9210	0.9125	0.9410	0.9530	0.9475	0.9210	0.9125
	AL	1.0718	1.0931	1.0857	0.9470	1.1462	0.9433	0.9984	0.9680	0.9470	1.1462
	SD	1.7330	1.2193	1.4689	0.4887	0.5367	1.2451	0.8760	1.0553	0.4887	0.5367
$n = 40$	Bias	-0.0079	0.0076	0.0020	-0.0056	0.0564	-0.0079	0.0076	0.0020	-0.0056	0.0564
	CP	0.9295	0.9285	0.9265	0.9265	0.9125	0.9380	0.9530	0.9445	0.9265	0.9125
	AL	0.9337	0.9501	0.9434	0.8389	1.0114	0.7989	0.8358	0.8140	0.8389	1.0114
	SD	1.7319	1.2283	1.4813	0.4603	0.5049	1.2395	0.8791	1.0601	0.4603	0.5049

TABLE 2: Performance of the point estimators and 95% CIs for response mean in simulation 1, $\varepsilon_i \sim N(0, 1 + 0.5X^2)$.

Model		OEL					MEL				
		$\hat{\mu}_{ipw}$	$\hat{\mu}_{mi}$	$\hat{\mu}_{aipw}$	$\bar{Y}$	$\hat{\mu}_{cc}$	$\hat{\mu}_{ipw}$	$\hat{\mu}_{mi}$	$\hat{\mu}_{aipw}$	$\bar{Y}$	$\hat{\mu}_{cc}$
$n = 20$	Bias	-0.0103	0.0146	0.0085	0.0011	0.0588	-0.0103	0.0146	0.0085	0.0011	0.0588
	CP	0.9080	0.9155	0.9100	0.9175	0.9210	0.9280	0.9390	0.9365	0.9175	0.9210
	AL	1.5912	1.6347	1.6177	1.4339	1.7487	1.4772	1.5965	1.5263	1.4339	1.7487
	SD	2.0256	1.4345	1.7641	0.6003	0.6605	1.4662	1.0385	1.2770	0.6003	0.6605
$n = 30$	Bias	-0.0101	0.0077	0.0025	-0.0076	0.0508	-0.0101	0.0077	0.0025	-0.0076	0.0508
	CP	0.9235	0.9275	0.9245	0.9145	0.9140	0.9400	0.9545	0.9440	0.9145	0.9140
	AL	1.3317	1.3684	1.3492	1.1524	1.4058	1.1797	1.2651	1.2083	1.1524	1.4058
	SD	2.0550	1.4687	1.8080	0.5385	0.5935	1.4764	1.0552	1.2990	0.5385	0.5935
$n = 40$	Bias	-0.0137	0.0024	-0.0041	-0.0092	0.0503	-0.0137	0.0024	-0.0041	-0.0092	0.0503
	CP	0.9250	0.9300	0.9290	0.9280	0.9160	0.9355	0.9540	0.9440	0.9280	0.9160
	AL	1.1636	1.1943	1.1765	1.0196	1.2421	1.0004	1.0637	1.0190	1.0196	1.2421
	SD	2.0554	1.4799	1.8286	0.5070	0.5589	1.4710	1.0592	1.3087	0.5070	0.5589

$(\hat{\mu} - 1.96\hat{\sigma}_{se}, \hat{\mu} + 1.96\hat{\sigma}_{se})$, with standard error $\hat{\sigma}_{se}$ obtained by the square roots of bootstrap variance estimators based on $n/4$ bootstrap replications, where n is the size of the sample.

From Tables 3 and 4, we can see the following.

- (1) Comparison for different sample sizes: when the sample sizes increase, estimators $\hat{\mu}_{ipw}$, $\hat{\mu}_{mi}$, and $\hat{\mu}_{aipw}$ have small biases, the coverage probability is big, the AL and the SD are small. So, the finite sample performance of the mean empirical likelihood is as that of the original empirical likelihood
- (2) Comparison for different estimation methods: estimators $\hat{\mu}_{ipw}$, $\hat{\mu}_{mi}$, and $\hat{\mu}_{aipw}$ based on MEL have similar point estimation compared with those based on OEL. In addition, estimators $\hat{\mu}_{ipw}$, $\hat{\mu}_{mi}$, and $\hat{\mu}_{aipw}$ based on MEL have bigger CP and smaller SD than those based on OEL. So, the ALs based on MEL are comparable, close to the method based on $\bar{Y}$
- (3) Comparison for different estimating equations: as to the bias, the estimator $\hat{\mu}_{aipw}$ is closer than $\hat{\mu}_{ipw}$ and $\hat{\mu}_{mi}$. The CP of the estimator $\hat{\mu}_{aipw}$ performs better than $\hat{\mu}_{ipw}$ and $\hat{\mu}_{mi}$ in many cases. So, the AL of the estimator $\hat{\mu}_{ipw}$ performs better than $\hat{\mu}_{aipw}$ and $\hat{\mu}_{mi}$ in many cases

- (4) Comparison for different error situations: the variation range of the bias estimated by OEL and MEL with $\varepsilon_i \sim N(0, 1 + 0.5X^2)$ is much smaller and more stable than the method with $\varepsilon_i \sim N(0, 1)$. But the standard deviation estimated by OEL and MEL with $\varepsilon_i \sim N(0, 1 + 0.5X^2)$ is close to the method with $\varepsilon_i \sim N(0, 1)$
- (5) Comparison for different correlation structures: when ρ is large, the biases based on MEL are not stable but are compared to $\hat{\mu}_{cc}$. But the CPs and the ALs still perform well, and the ALs are stable. when $\rho = 0$ and $\rho = 0.2$, the MEL method performs better than the OEL method

In a word, the performances of IPW, MI, and AIPW based on MEL precede those methods based on OEL, and AIPW based on MEL achieves the most accurate results.

4. Real Data Analysis

In this section, we apply the proposed method to the Boston Housing data to illustrate our proposed MEL for missing data. The data is taken from the UCI Irvine Machine Learning Repository. The distribution of the per capita crime rate (CRIM) by the town is unknown. The inference on

TABLE 3: Performance of the point estimators and 95% CIs for response mean in simulation 2, $\varepsilon_i \sim N(0, 1)$.

Model		OEL				MEL			
		$\hat{\mu}_{ipw}$	$\hat{\mu}_{mi}$	$\hat{\mu}_{aipw}$	$\hat{\mu}_{ipw}$	$\hat{\mu}_{mi}$	$\hat{\mu}_{aipw}$	$\bar{Y}$	$\hat{\mu}_{cc}$
$\rho = 0$									
M1									
n = 20	Bias	0.0808	0.0735	0.0569	0.0808	0.0735	0.0569	-0.0057	0.2179
	CP	0.8830	0.8810	0.8730	0.9060	0.9160	0.9040	0.8600	0.8460
	AL	2.2307	2.2362	2.2253	2.0756	2.1483	2.0939	2.0802	2.5970
	SD	2.8128	2.1589	2.4451	2.0376	1.5667	1.7734	0.7145	0.7947
	Bias	0.0802	0.0593	0.0465	0.0802	0.0593	0.0465	0.0005	0.2186
n = 40	CP	0.9250	0.9220	0.9200	0.9360	0.9390	0.9340	0.8720	0.8380
	AL	1.6084	1.6041	1.5973	1.3785	1.3956	1.3773	1.3959	1.7330
	SD	2.8762	2.2972	2.5172	2.0585	1.6442	1.8017	0.5858	0.6519
M2									
n = 20	Bias	0.1208	0.1080	0.0809	0.1208	0.1080	0.0809	-0.0057	0.3595
	CP	0.8960	0.8810	0.8860	0.9210	0.9120	0.9140	0.8600	0.7970
	AL	2.2699	2.2669	2.2638	2.1130	2.1598	2.1265	2.0802	2.4904
	SD	2.7870	2.2731	2.4981	2.0190	1.6488	1.8112	0.7145	0.7792
	Bias	0.0904	0.0613	0.0334	0.0903	0.0613	0.0334	0.0005	0.3539
n = 40	CP	0.9360	0.9350	0.9290	0.9460	0.9460	0.9420	0.8720	0.7700
	AL	1.6311	1.6268	1.6221	1.3990	1.4102	1.3978	1.3959	1.6898
	SD	2.8535	2.3950	2.5621	2.0423	1.7142	1.8338	0.5858	0.6434
M3									
n = 20	Bias	0.1151	0.1058	0.0827	0.1151	0.1058	0.0827	-0.0057	0.3179
	CP	0.8880	0.8880	0.8810	0.9080	0.9110	0.9090	0.8600	0.8320
	AL	2.2304	2.2318	2.2233	2.0755	2.1359	2.0897	2.0802	2.5149
	SD	2.7751	2.1871	2.4485	2.0104	1.5867	1.7754	0.7145	0.7820
	Bias	0.1063	0.0777	0.0574	0.1063	0.0777	0.0574	0.0005	0.3119
n = 40	CP	0.9170	0.9200	0.9200	0.9340	0.9350	0.9350	0.8720	0.7810
	AL	1.6066	1.6015	1.5959	1.3774	1.3913	1.3756	1.3959	1.6988
	SD	2.8470	2.3150	2.5181	2.0376	1.6570	1.8023	0.5858	0.6447
$\rho = 0.2$									
M1									
n = 20	Bias	0.1154	0.1044	0.0841	0.1154	0.1044	0.0841	0.0054	0.3065
	CP	0.8710	0.8790	0.8700	0.8990	0.9070	0.9110	0.8670	0.8330
	AL	2.5572	2.5494	2.5434	2.3828	2.4450	2.3943	2.3166	2.9036
	SD	3.2147	2.4859	2.8035	2.3303	1.8076	2.0366	0.7536	0.8401
	Bias	0.0892	0.0646	0.0481	0.0892	0.0646	0.0481	-0.0031	0.2807
n = 40	CP	0.9140	0.9080	0.9120	0.9330	0.9250	0.9250	0.8530	0.8190
	AL	1.8363	1.8334	1.8259	1.5728	1.5948	1.5742	1.5825	2.0088
	SD	3.3133	2.6278	2.8788	2.3716	1.8811	2.0606	0.6238	0.7026
M2									
n = 20	Bias	0.1713	0.1592	0.1247	0.1713	0.1592	0.1247	0.0054	0.4961
	CP	0.8980	0.8990	0.9020	0.9200	0.9230	0.9300	0.8670	0.7940
	AL	2.5756	2.5736	2.5706	2.3966	2.4466	2.4143	2.3166	2.7918
	SD	3.1657	2.6116	2.8377	2.2943	1.8953	2.0584	0.7536	0.8257
	Bias	0.0969	0.0686	0.0330	0.0969	0.0686	0.0330	-0.0031	0.4320
n = 40	CP	0.9270	0.9240	0.9300	0.9390	0.9450	0.9430	0.8530	0.7840
	AL	1.8630	1.8599	1.8546	1.5974	1.6114	1.5978	1.5825	1.9375
	SD	3.2785	2.7502	2.9305	2.3467	1.9686	2.0976	0.6238	0.6895
M3									
n = 20	Bias	0.1617	0.1441	0.1180	0.1617	0.1441	0.1180	0.0054	0.4538
	CP	0.8770	0.8730	0.8710	0.9050	0.9020	0.9050	0.8670	0.8020
	AL	2.5628	2.5578	2.5536	2.3868	2.4466	2.4015	2.3166	2.8692
	SD	3.1998	2.5164	2.8125	2.3196	1.8284	2.0416	0.7536	0.8361
	Bias	0.1357	0.0998	0.0731	0.1357	0.0998	0.0731	-0.0031	0.4151
n = 40	CP	0.9130	0.9110	0.9140	0.9210	0.9290	0.9310	0.8530	0.7560
	AL	1.8326	1.8268	1.8218	1.5709	1.5859	1.5705	1.5825	1.9471
	SD	3.2652	2.6553	2.8739	2.3372	1.9007	2.0571	0.6238	0.6908

TABLE 3: Continued.

Model		OEL				MEL		$\bar{Y}$	$\hat{\mu}_{cc}$
		$\hat{\mu}_{ipw}$	$\hat{\mu}_{mi}$	$\hat{\mu}_{aipw}$	$\hat{\mu}_{ipw}$	$\hat{\mu}_{mi}$	$\hat{\mu}_{aipw}$		
$\rho = 0.5$									
<i>M1</i>									
$n = 20$	Bias	0.1535	0.1411	0.1149	0.1535	0.1411	0.1149	0.0089	0.5182
	CP	0.8710	0.8710	0.8690	0.9040	0.8960	0.9030	0.8600	0.8020
	AL	3.2116	3.1963	3.1949	2.9926	3.0561	3.0066	2.9522	3.6533
	SD	4.0680	3.1548	3.5297	2.9527	2.2992	2.5686	0.8511	0.9427
$n = 40$	Bias	0.1413	0.1051	0.0787	0.1413	0.1051	0.0787	-0.0012	0.4787
	CP	0.9010	0.9070	0.9060	0.9160	0.9250	0.9210	0.8540	0.7730
	AL	2.3034	2.2967	2.2896	1.9729	1.9955	1.9740	2.0039	2.5153
	SD	4.2003	3.3200	3.6119	3.0074	2.3778	2.5865	0.7018	0.7849
<i>M2</i>									
$n = 20$	Bias	0.2056	0.1901	0.1481	0.2056	0.1901	0.1481	0.0089	0.7143
	CP	0.8950	0.8950	0.8990	0.9220	0.9200	0.9280	0.8600	0.7810
	AL	3.2669	3.2575	3.2575	3.0365	3.0899	3.0546	2.9522	3.5512
	SD	4.0194	3.3494	3.6082	2.9177	2.4359	2.6216	0.8511	0.9314
$n = 40$	Bias	0.1260	0.0965	0.0469	0.1260	0.0965	0.0469	-0.0012	0.6236
	CP	0.9330	0.9300	0.9360	0.9430	0.9400	0.9420	0.8540	0.7430
	AL	2.3612	2.3569	2.3515	2.0244	2.0398	2.0258	2.0039	2.4226
	SD	4.1723	3.5164	3.7187	2.9872	2.5179	2.6625	0.7018	0.7707
<i>M3</i>									
$n = 20$	Bias	0.3124	0.2802	0.2289	0.3124	0.2802	0.2289	0.0151	1.2031
	CP	0.8730	0.8510	0.8750	0.9050	0.8830	0.9060	0.8690	0.7220
	AL	4.1418	4.0918	4.1109	3.8600	3.8767	3.8564	3.8603	4.6516
	SD	5.2121	4.1575	4.5768	3.7932	3.0398	3.3380	0.9735	1.0647
$n = 40$	Bias	0.1927	0.1376	0.0925	0.1927	0.1376	0.0925	-0.0012	0.6938
	CP	0.9120	0.9070	0.9170	0.9220	0.9230	0.9290	0.8540	0.6960
	AL	2.3148	2.3015	2.2977	1.9848	1.9965	1.9808	2.0039	2.4544
	SD	4.1719	3.3648	3.6280	2.9872	2.4098	2.5979	0.7018	0.7748
$\rho = 0.8$									
<i>M1</i>									
$n = 20$	Bias	0.2208	0.2108	0.1787	0.2208	0.2108	0.1787	0.0151	0.8554
	CP	0.8630	0.8510	0.8630	0.8960	0.8940	0.9030	0.8690	0.7790
	AL	4.1478	4.1197	4.1233	3.8681	3.9367	3.8769	3.8603	4.7258
	SD	5.2678	4.1058	4.5875	3.8329	3.0062	3.3501	0.9735	1.0731
$n = 40$	Bias	0.2078	0.1658	0.1202	0.2078	0.1658	0.1202	0.0027	0.8162
	CP	0.9000	0.9060	0.9130	0.9200	0.9190	0.9250	0.8560	0.7350
	AL	2.9851	2.9730	2.9672	2.5576	2.5821	2.5593	2.6054	3.2279
	SD	5.4854	4.3172	4.6855	3.9303	3.0954	3.3581	0.8000	0.8886
<i>M2</i>									
$n = 20$	Bias	0.2486	0.2389	0.1806	0.2486	0.2389	0.1806	0.0151	1.0812
	CP	0.8970	0.8810	0.8970	0.9240	0.9090	0.9220	0.8690	0.7530
	AL	4.2464	4.2105	4.2197	3.9523	3.9785	3.9607	3.8603	4.5830
	SD	5.2769	4.4156	4.7184	3.8404	3.2228	3.4385	0.9735	1.0572
$n = 40$	Bias	0.1611	0.1431	0.0709	0.1611	0.1431	0.0709	0.0027	0.9679
	CP	0.9300	0.9280	0.9330	0.9470	0.9360	0.9500	0.8560	0.6910
	AL	3.0743	3.0644	3.0606	2.6322	2.6501	2.6364	2.6054	3.0973
	SD	5.4867	4.6103	4.8460	3.9307	3.3039	3.4719	0.8000	0.8721
<i>M3</i>									
$n = 20$	Bias	0.3124	0.2802	0.2289	0.3124	0.2802	0.2289	0.0151	1.2031
	CP	0.8730	0.8510	0.8750	0.9050	0.8830	0.9060	0.8690	0.7220
	AL	4.1418	4.0918	4.1109	3.8600	3.8767	3.8564	3.8603	4.6516
	SD	5.2121	4.1575	4.5768	3.7932	3.0398	3.3380	0.9735	1.0647
$n = 40$	Bias	0.2910	0.2193	0.1472	0.2910	0.2193	0.1472	0.0027	1.1758
	CP	0.9040	0.9000	0.9110	0.9200	0.9220	0.9370	0.8560	0.6100
	AL	3.0001	2.9723	2.9738	2.5739	2.5768	2.5644	2.6054	3.1358
	SD	5.4574	4.3697	4.6988	3.9111	3.1328	3.3671	0.8000	0.8764

TABLE 4: Performance of the point estimators and 95% CIs for response mean in simulation 2, $\varepsilon_i \sim N(0, 1 + 0.5X_{i1}^2)$.

Model		OEL				MEL		$\bar{Y}$	$\hat{\mu}_{cc}$
		$\hat{\mu}_{ipw}$	$\hat{\mu}_{mi}$	$\hat{\mu}_{aipw}$	$\hat{\mu}_{ipw}$	$\hat{\mu}_{mi}$	$\hat{\mu}_{aipw}$		
$\rho = 0$									
M1									
$n = 20$	Bias	0.0980	0.0955	0.0804	0.0980	0.0955	0.0804	−0.0053	0.2177
	CP	0.8900	0.8810	0.8840	0.9100	0.9120	0.9150	0.8530	0.8350
	AL	2.3574	2.3634	2.3540	2.1989	2.2797	2.2196	2.1519	2.6703
	SD	2.9416	2.2382	2.5761	2.1315	1.6258	1.8695	0.7264	0.8057
$n = 40$	Bias	0.0606	0.0401	0.0268	0.0606	0.0401	0.0268	0.0019	0.2058
	CP	0.9230	0.9210	0.9190	0.9330	0.9370	0.9330	0.8680	0.8320
	AL	1.6724	1.6690	1.6613	1.4337	1.4539	1.4328	1.4410	1.8057
	SD	2.9763	2.3670	2.6156	2.1302	1.6942	1.8721	0.5948	0.6646
M2									
$n = 20$	Bias	0.1227	0.1117	0.0856	0.1227	0.1117	0.0856	−0.0053	0.3586
	CP	0.8790	0.8790	0.8810	0.9020	0.9150	0.9060	0.8530	0.8150
	AL	2.3549	2.3568	2.3487	2.1940	2.2495	2.2070	2.1519	2.5743
	SD	2.8880	2.3479	2.5898	2.0923	1.7041	1.8787	0.7264	0.7921
$n = 80$	Bias	0.0811	0.0508	0.0243	0.0811	0.0508	0.0243	0.0019	0.3421
	CP	0.9360	0.9310	0.9350	0.9440	0.9460	0.9470	0.8680	0.7940
	AL	1.6970	1.6932	1.6876	1.4564	1.4697	1.4549	1.4410	1.7536
	SD	2.9564	2.4705	2.6611	2.1159	1.7682	1.9046	0.5948	0.6549
M3									
$n = 20$	Bias	0.1392	0.1266	0.1050	0.1392	0.1266	0.1050	−0.0053	0.3263
	CP	0.8780	0.8790	0.8810	0.9110	0.9150	0.9160	0.8530	0.8290
	AL	2.3407	2.3431	2.3361	2.1820	2.2460	2.1991	2.1519	2.5981
	SD	2.8894	2.2687	2.5650	2.0932	1.6466	1.8604	0.7264	0.7947
$n = 80$	Bias	0.0918	0.0650	0.0437	0.0918	0.0650	0.0437	0.0019	0.3066
	CP	0.9290	0.9280	0.9320	0.9390	0.9460	0.9430	0.8680	0.8030
	AL	1.6750	1.6684	1.6629	1.4372	1.4516	1.4344	1.4410	1.7434
	SD	2.9589	2.3867	2.6183	2.1177	1.7083	1.8740	0.5948	0.6531
$\rho = 0.2$									
M1									
$n = 20$	Bias	0.0991	0.0980	0.0789	0.0991	0.0980	0.0789	−0.0057	0.2874
	CP	0.8750	0.8730	0.8770	0.8970	0.9150	0.9040	0.8580	0.8330
	AL	2.6330	2.6399	2.6297	2.4551	2.5450	2.4793	2.3812	2.9765
	SD	3.3056	2.5042	2.8798	2.3965	1.8204	2.0915	0.7651	0.8508
$n = 40$	Bias	0.0666	0.0452	0.0262	0.0666	0.0452	0.0262	−0.0068	0.2749
	CP	0.9020	0.9150	0.9110	0.9100	0.9350	0.9290	0.8580	0.8030
	AL	1.9021	1.8972	1.8912	1.6299	1.6517	1.6316	1.6631	2.0389
	SD	3.4267	2.6980	2.9757	2.4527	1.9313	2.1300	0.6389	0.7060
M2									
$n = 20$	Bias	0.1265	0.1133	0.0843	0.1265	0.1133	0.0843	−0.0057	0.4260
	CP	0.8950	0.9000	0.9000	0.9190	0.9270	0.9250	0.8580	0.7880
	AL	2.6613	2.6589	2.6522	2.4787	2.5309	2.4908	2.3812	2.8391
	SD	3.2597	2.6807	2.9323	2.3636	1.9468	2.1283	0.7651	0.8321
$n = 40$	Bias	0.0892	0.0567	0.0244	0.0892	0.0567	0.0244	−0.0068	0.4145
	CP	0.9230	0.9230	0.9190	0.9340	0.9350	0.9360	0.8580	0.7650
	AL	1.9339	1.9289	1.9246	1.6591	1.6722	1.6590	1.6631	1.9615
	SD	3.3837	2.8391	3.0365	2.4220	2.0323	2.1735	0.6389	0.6934
M3									
$n = 20$	Bias	0.1646	0.1477	0.1237	0.1646	0.1477	0.1237	−0.0057	0.4220
	CP	0.8830	0.8730	0.8790	0.9020	0.9100	0.9070	0.8580	0.8010
	AL	2.5971	2.5984	2.5888	2.4225	2.4859	2.4314	2.3812	2.8912
	SD	3.2057	2.5433	2.8497	2.3243	1.8481	2.0690	0.7651	0.8394
$n = 40$	Bias	0.1091	0.0742	0.0478	0.1091	0.0742	0.0478	−0.0068	0.3910
	CP	0.9110	0.9110	0.9140	0.9270	0.9400	0.9310	0.8580	0.7740
	AL	1.9105	1.9025	1.8980	1.6388	1.6541	1.6372	1.6631	2.0048
	SD	3.3896	2.7397	2.9884	2.4262	1.9612	2.1392	0.6389	0.7006

TABLE 4: Continued.

Model		OEL				MEL		$\bar{Y}$	$\hat{\mu}_{cc}$
		$\hat{\mu}_{ipw}$	$\hat{\mu}_{mi}$	$\hat{\mu}_{aipw}$	$\hat{\mu}_{ipw}$	$\hat{\mu}_{mi}$	$\hat{\mu}_{aipw}$		
$\rho = 0.5$									
M1									
$n = 20$	Bias	0.1580	0.1465	0.1176	0.1580	0.1465	0.1176	-0.0058	0.4920
	CP	0.8800	0.8690	0.8780	0.9070	0.9030	0.9090	0.8560	0.8150
	AL	3.2556	3.2475	3.2379	3.0391	3.1207	3.0446	2.9838	3.7143
	SD	4.0892	3.1324	3.5685	2.9687	2.2847	2.5988	0.8565	0.9512
$n = 40$	Bias	0.1102	0.0826	0.0527	0.1102	0.0826	0.0527	-0.0046	0.4445
	CP	0.9180	0.9120	0.9140	0.9290	0.9270	0.9290	0.8630	0.7860
	AL	2.3729	2.3659	2.3596	2.0330	2.0576	2.0355	2.0905	2.5668
	SD	4.3127	3.3914	3.7184	3.0879	2.4292	2.6629	0.7166	0.7921
M2									
$n = 20$	Bias	0.1523	0.1363	0.0973	0.1523	0.1363	0.0973	-0.0058	0.6213
	CP	0.8960	0.8920	0.8910	0.9210	0.9240	0.9310	0.8560	0.7880
	AL	3.3218	3.3080	3.3102	3.0877	3.1327	3.1067	2.9838	3.5378
	SD	4.0869	3.4004	3.6695	2.9676	2.4738	2.6672	0.8565	0.9286
$n = 40$	Bias	0.1219	0.0935	0.0426	0.1219	0.0935	0.0426	-0.0046	0.6363
	CP	0.9200	0.9200	0.9270	0.9310	0.9320	0.9320	0.8630	0.7240
	AL	2.4278	2.4195	2.4158	2.0826	2.0947	2.0818	2.0905	2.4501
	SD	4.2839	3.6014	3.8182	3.0675	2.5793	2.7343	0.7166	0.7754
M3									
$n = 20$	Bias	0.2194	0.2007	0.1642	0.2194	0.2007	0.1642	-0.0058	0.6914
	CP	0.8700	0.8560	0.8690	0.8990	0.8900	0.9030	0.8560	0.7650
	AL	3.2062	3.1934	3.1938	2.9879	3.0523	3.0051	2.9838	3.5663
	SD	3.9945	3.1661	3.5255	2.9018	2.3080	2.5661	0.8565	0.9324
$n = 40$	Bias	0.1856	0.1361	0.0911	0.1856	0.1361	0.0911	-0.0046	0.6684
	CP	0.9200	0.9100	0.9180	0.9400	0.9330	0.9350	0.8630	0.7210
	AL	2.3662	2.3524	2.3495	2.0292	2.0416	2.0258	2.0905	2.5127
	SD	4.2445	3.4271	3.7071	3.0391	2.4546	2.6548	0.7166	0.7847
$\rho = 0.8$									
M1									
$n = 20$	Bias	0.2201	0.2048	0.1692	0.2201	0.2048	0.1692	0.0042	0.8355
	CP	0.8470	0.8410	0.8440	0.8740	0.8770	0.8840	0.8400	0.7960
	AL	4.2359	4.2035	4.2082	3.9467	4.0136	3.9594	4.0717	5.0903
	SD	5.3800	4.1687	4.6843	3.9171	3.0545	3.4230	0.9983	1.1108
$n = 40$	Bias	0.1813	0.1383	0.0916	0.1813	0.1383	0.0916	-0.0020	0.7745
	CP	0.9200	0.9100	0.9210	0.9310	0.9340	0.9380	0.8580	0.7410
	AL	3.0360	3.0211	3.0169	2.6021	2.6244	2.6029	2.7111	3.2838
	SD	5.5618	4.3777	4.7587	3.9845	3.1381	3.4099	0.8163	0.8960
M2									
$n = 20$	Bias	0.1914	0.1790	0.1246	0.1914	0.1790	0.1246	-0.0093	0.9730
	CP	0.9070	0.8890	0.9110	0.9280	0.9230	0.9310	0.8640	0.7640
	AL	4.2705	4.2472	4.2587	3.9727	4.0115	3.9929	3.8611	4.5488
	SD	5.3181	4.4389	4.7489	3.8693	3.2383	3.4595	0.9743	1.0532
$n = 40$	Bias	0.1519	0.1380	0.0644	0.1519	0.1380	0.0644	-0.0020	0.9899
	CP	0.9150	0.9160	0.9200	0.9330	0.9310	0.9390	0.8580	0.6760
	AL	3.1334	3.1201	3.1183	2.6876	2.6998	2.6878	2.7111	3.1629
	SD	5.6144	4.6827	4.9336	4.0228	3.3566	3.5354	0.8163	0.8809
M3									
$n = 20$	Bias	0.2897	0.2575	0.2094	0.2897	0.2575	0.2094	-0.0093	1.1480
	CP	0.8500	0.8490	0.8640	0.8830	0.8710	0.8920	0.8640	0.7380
	AL	4.1124	4.0760	4.0890	3.8351	3.8662	3.8430	3.8611	4.5670
	SD	5.1916	4.1105	4.5624	3.7832	3.0098	3.3325	0.9743	1.0546
$n = 40$	Bias	0.2669	0.2032	0.1309	0.2669	0.2032	0.1309	-0.0020	1.1346
	CP	0.9030	0.9050	0.9190	0.9230	0.9210	0.9340	0.8580	0.6430
	AL	3.0494	3.0194	3.0215	2.6178	2.6196	2.6073	2.7111	3.1736
	SD	5.5468	4.4192	4.7703	3.9751	3.1687	3.4189	0.8163	0.8827

TABLE 5: Performance of estimation for response mean in the Boston data (rep = 1000, $b = 20$, $n = 20$).

Bias method	IPW MEL	3.537 OEL	MI MEL	3.548 OEL	AIPW MEL	3.538 OEL	Y: 3.609	CC: 3.899
CP	0.885	0.855	0.88	0.855	0.887	0.859	0.773	0.807
AL	6.2903	6.6173	6.3841	6.6196	6.3362	6.6501	5.8252	6.3198
SD	5.6104	7.6612	5.1066	6.9549	5.4856	7.4860	1.1551	1.2012

TABLE 6: Performance of estimation for response mean in the Boston data (rep = 1000, $b = 20$, $n = 40$).

Bias method	IPW MEL	3.501 OEL	MI MEL	3.501 OEL	AIPW MEL	3.504 OEL	Y: 3.5898	CC: 3.8859
CP	0.891	0.867	0.893	0.867	0.892	0.868	0.843	0.817
AL	4.4330	4.9411	4.4381	4.9429	4.4349	4.9419	4.8270	4.4573
SD	5.4718	7.6169	5.4325	7.5619	5.4518	7.5889	1.0702	1.0291

TABLE 7: Performance of estimation for response mean in the Boston data (rep = 1000, $b = 20$, $n = 80$).

Bias Method	IPW MEL	3.595 OEL	MI MEL	3.600 OEL	AIPW MEL	3.599 OEL	Y: 3.649	CC: 3.956
CP	0.918	0.904	0.918	0.905	0.918	0.903	0.85	0.88
AL	3.1744	3.7868	3.1785	3.7889	3.1766	3.7881	3.4933	3.7886
SD	5.8355	8.1944	5.7943	8.1365	5.8158	8.1666	0.9216	0.9600

CRIM distribution based on the full data has even been analyzed by Liang et al. [15]. The dataset consists of 506 observations and 14 variates, with CRIM being the response variate and the other 11 being the covariates. We are interested in the mean of CRIM and its confidence region. We set the unconditional rates of missing data between 5% and 15%, and all real data analysis results are based on 1000 replications and the sample size is $n = 20$, $n = 40$, and $n = 80$. Tables 5–7 show the biases and standard deviations (SDs) of the point estimators, the average lengths (ALs), and the coverage probabilities (CPs) of confidence intervals (CIs) at the nominal level 95%.

In Tables 5–7, it is observed that when the sample sizes increase, the coverage probabilities of the estimators $\hat{\mu}_{ipw}$, $\hat{\mu}_{mi}$, and $\hat{\mu}_{aipw}$ based on the MEL method are bigger than those of the AL and the SD. In particular, the confidence regions of the response mean are close to the confidence region of the full data. So, the finite sample performance of mean empirical likelihood is better than that of the original empirical likelihood.

5. Conclusions

This paper proposed a new empirical likelihood for inference on response mean with a small sample and multidimensional variate using the mean empirical likelihood. Its large sample properties were proved with different bias-corrected estimating equations. This new method outperforms the original empirical likelihood methods.

On the basis of IPW, MI, and AIPW, the new method uses a pairwise-mean dataset to obtain its advantages. Compared with the original EL method, this method can obtain an approximate point estimator and better CPs, ALs, and SDs. Compared with the other EL inference methods, this method has two characteristics: one is that the size of the sample is much larger than dimensions, and the other is that the number of samples is large. In addition, AIPW based on MEL achieves

the most accurate results. Theoretically, when the dimension of the covariate is increasing, the performance of response mean inference based EL is bad. The method of Liang et al. [15] can solve this problem. Meanwhile, it has superior performance when the size of the sample is small. Therefore, we recommend that three bias-corrected models be carefully constructed and used to infer the confidence region of response mean with missing at random. We will further study the MEL method for categorical data with random missing covariates.

Appendix

- (C1) The propensity $\pi(X)$, the X -density function $f(X)$, and $m(X)$ all have continuous and bounded partial derivatives with respect to X up to order r with $r \geq 2$ and $r > d/2$, where r is the order of the kernel K , and the propensity $\pi(X)$ is bounded away from 0 to 1.
- (C2) $E(Y^2)$ is finite.
- (C3) As $n \rightarrow \infty$, $nh^{2d} \rightarrow \infty$, $nh^d/\log n \rightarrow \infty$, $nh^{2m} \rightarrow \infty$.
- (C4) $\text{Cov}(\phi_l) = \Sigma$ exists, where $\text{rank}(\Sigma) = d$.
- (C5) When $X \sim N(\mu, \sigma^2)$, then $(X - \mu)/\sigma \sim N(0, 1)$.

Lemma A.1. Satisfying situations (C1)–(C4),

- (i) $N^{-1/2} \sum_{k=1}^N \hat{V}_{lk} \rightarrow^L N(0, \sigma^2)$
- (ii) $(1/N) \sum_{k=1}^N \hat{\phi}_{li}^2(\mu_0) \rightarrow (1/2)\sigma_l^2$

Proof

- (i) According to Lemma A.1
- (ii) in Liang et al. [15], noticing

$$\begin{aligned} \frac{1}{N} \sum_{k=1}^N \widehat{V}_{lk}(\mu_0) &= \frac{1}{2N} \left(\sum_{i=1}^n \sum_{j=1}^n \frac{\widehat{\phi}_{li} + \widehat{\phi}_{lj}}{2} + \sum_{i=1}^n \widehat{\phi}_{li} \right) \\ &= \frac{n+1}{2N} \sum_{i=1}^n \widehat{\phi}_{li} = \bar{\phi}_{ln}, \end{aligned} \quad (\text{A.1})$$

and following the assumption $\text{Cov}(V_k) = \Sigma$, we can obtain $\bar{\phi}_{ln} = (1/N) \sum_{i=1}^n \widehat{\phi}_{li} = O_p(n^{-1/2})$; therefore,

$$\frac{1}{N} \sum_{k=1}^N \widehat{V}_{lk} = O_p(n^{-1/2}). \quad (\text{A.2})$$

Then, we prove that (i) holds.

(ii) According to Lemma A.1 (iv) in Liang et al. [15], we notice that

$$\begin{aligned} \frac{1}{N} \sum_{k=1}^N \widehat{V}_{lk}^2(\mu_0) &= \frac{1}{N} \sum_{k=1}^N \|\widehat{V}_{lk}\|^2 \\ &= \frac{n+2}{2(n+1)} \left(\frac{1}{n} \sum_{i=1}^n \widehat{\phi}_{li}^T \widehat{\phi}_{li} \right) + \frac{1}{2(n+1)} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \widehat{\phi}_{li}^T \right) \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \widehat{\phi}_{li} \right) \\ &= \frac{n+2}{2(n+1)} \left(\frac{1}{n} \sum_{i=1}^n \widehat{\phi}_{li}^2 \right) + O_p(1) \\ &= \frac{n+2}{2(n+1)} \left(\frac{1}{n} \sum_{i=1}^n \{M_i Y_i + (1 - \delta_i) m(X_i) - \mu_0\}^2 \right) + O_p(1) \\ &= \frac{n+2}{2(n+1)} \sigma_l^2 + O_p(1). \end{aligned} \quad (\text{A.3})$$

Therefore, $(1/N) \sum_{k=1}^N \widehat{\phi}_{li}^2(\mu_0) \longrightarrow (1/2) \sigma_l^2$. $\square$

Proof of Theorem 1. Similar to the proof of Theorem 1 in Wang and Deng [9], we can prove Theorem 1. $\square$

Proof of Theorem 2. For $\widehat{R}_l(\mu_0)$, $l = 1, 2, 3$, using the arguments in the Wang and Rao [4], we have

$$\begin{aligned} \widehat{R}_l^M(\mu_0) &= 2 \sum_{k=1}^N \log\{1 + \lambda_l \widehat{V}_{lk}(\mu_0)\} \\ &= \left\{ \frac{1}{\sqrt{N}} \sum_{k=1}^N \widehat{V}_{lk}(\mu_0) \right\} \left\{ \frac{1}{N} \sum_{k=1}^N \widehat{V}_{lk}^2(\mu_0) \right\}^{-1} \left\{ \frac{1}{\sqrt{N}} \sum_{k=1}^N \widehat{V}_{lk}(\mu_0) \right\} + O_p(1). \end{aligned} \quad (\text{A.4})$$

Using Theorem 2 and Lemma (i), we have $(1/\sqrt{N} \sigma) \sum_{k=1}^N V_{lk}(\mu_0) \xrightarrow{L} N(0, 1)$. Therefore, we have

$$\widehat{R}_l^M(\mu_0) = 2 \frac{\sigma^2}{\sigma_l^2} \left\{ \frac{1}{\sqrt{N} \sigma} \sum_{k=1}^N \widehat{V}_{lk}(\mu_0) \right\}^2 + O_p(1) \xrightarrow{L} 2\vartheta_l \chi_1^2, \quad (\text{A.5})$$

where $\vartheta_l = \sigma^2/\sigma_l^2$, $l = 1, 2$, and $\vartheta_3 = 1$. $\square$

Proof of Theorem 3. According to Lemma 11.1 in Owen [3], with probability tending to 1, 0 is inside the convex hull of

V_{lk} , $k = 1, 2, \dots, n$. By using the Lagrange multiplier, we have

$$p_k = \frac{1}{N(1 + \lambda_l' V_{lk})} > 0, \quad (\text{A.6})$$

where λ satisfies the equation $\sum p_k W_k = 1$. Then, applying Lemma A.1 in Liang et al. [15], we have

$$\|\lambda\| = O_p(n^{-1/2}). \quad (\text{A.7})$$

On the other hand, with the following equation [5, 16]:

$$0 = \frac{1}{N} \sum_{k=1}^N \frac{V_{lk}}{1 + \lambda_l' V_{lk}} = \frac{1}{N} \sum_{k=1}^N V_{lk} - \frac{1}{N} \left(\sum_{k=1}^N V_{lk} V_{lk}' \right) \lambda_l + \frac{1}{N} \sum_{k=1}^N \frac{V_{lk} (\lambda_l' V_{lk})^2}{1 + \lambda_l' V_{lk}}, \quad (\text{A.8})$$

we get

$$\lambda_l = \left(\sum_{k=1}^N V_{lk} V'_{lk} \right)^{-1} \left(\sum_{k=1}^N V_{lk} \right) + O_p(n^{-1/2}). \quad (\text{A.9})$$

Using Taylor's expansion, we can get

$$\begin{aligned} L_l^M(\mu_0) &= \frac{2}{n+1} \sum_{k=1}^N \log(1 + \lambda'_l V_{lk}) \\ &= \frac{2}{n+1} \sum_{k=1}^N \left(\lambda'_l V_{lk} - \frac{1}{2} (\lambda'_l V_{lk})^2 \right) + O_p(n). \end{aligned} \quad (\text{A.10})$$

Substituting λ into (A.10), we obtain

$$\begin{aligned} L_l^M(\mu_0) &= \frac{1}{n+1} \left(\sum_{k=1}^N V_{lk} \right) \left(\sum_{k=1}^N V_{lk} V'_{lk} \right)^{-1} \left(\sum_{k=1}^N V_{lk} \right) + O_p(1) \\ &= \frac{1}{n+1} (N \bar{V}_{\ln}) \left(N \cdot \frac{1}{2} \sum \right)^{-1} (N \bar{V}_{\ln}) + O_p(1) \\ &= \frac{2N}{n+1} \bar{V}'_{\ln} \sum \bar{V}_{\ln} + O_p(1) \\ &= n \bar{V}'_{\ln} \sum \bar{V}_{\ln} + O_p(1) \longrightarrow \chi_d^2, \text{ in dist.} \end{aligned} \quad (\text{A.11})$$

□

Data Availability

The Boston Housing data used in the real data analysis comes from an open-source database UCI Irvine Machine Learning Repository (<http://archive.ics.uci.edu/ml>).

Conflicts of Interest

The authors declare that there are no conflicts of interest in the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant no. 71963008) and the support program of the Guangxi China Science Foundation (grant no. 2018GXNSFAA294131).

References

- [1] A. B. Owen, "Empirical likelihood ratio confidence intervals for a single functional," *Biometrika*, vol. 75, no. 2, pp. 237–249, 1988.
- [2] J. Qin and J. Lawless, "Empirical likelihood and general estimating equations," *The Annals of Statistics*, vol. 22, no. 1, 1994.
- [3] A. B. Owen, *Empirical Likelihood*, Chapman and Hall/CRC Press, New York, NY, USA, 2001.
- [4] Q. Wang and J. N. K. Rao, "Empirical likelihood-based inference under imputation for missing response data," *The Annals of Statistics*, vol. 30, no. 3, pp. 896–924, 2002.
- [5] H. Liang, S. Wang, and R. J. Carroll, "Partially linear models with missing response variables and error-prone covariates," *Biometrika*, vol. 94, no. 1, pp. 185–198, 2007.
- [6] W. Stute, L. Xue, and L. Zhu, "Empirical likelihood inference in nonlinear errors-in-covariates models with validation data," *Journal of the American Statistical Association*, vol. 102, pp. 332–346, 2007.
- [7] Z. Guan, D. H. Y. Leung, and J. Qin, "Semiparametric maximum likelihood inference for nonignorable nonresponse with callbacks," *Scandinavian Journal of Statistics*, vol. 45, no. 4, pp. 962–984, 2018.
- [8] Y. Xie and B. Zhang, "Constrained empirical-likelihood confidence regions in nonignorable covariate-missing data problems," *Statistics in Medicine*, vol. 38, no. 3, pp. 452–479, 2019.
- [9] L. Wang and G. Deng, "Dimension-reduced empirical likelihood inference for response mean with data missing at random," *Journal of Nonparametric Statistics*, vol. 29, no. 3, pp. 1–21, 2017.
- [10] P. Hall and B. L. Scala, "Methodology and algorithms of empirical likelihood," *International Statistical Review/Revue Internationale de Statistique*, vol. 58, no. 2, pp. 109–127, 1990.
- [11] T. Diccio, P. Hall, and J. Romano, "Empirical likelihood is bartlett-correctable," *The Annals of Statistics*, vol. 19, no. 2, pp. 1053–1061, 1991.
- [12] M. Tsao, "Bounds on coverage probabilities of the empirical likelihood ratio confidence regions," *The Annals of Statistics*, vol. 32, no. 3, pp. 1215–1221, 2004.
- [13] J. Chen, A. M. Variyath, and B. Abraham, "Adjusted empirical likelihood and its properties," *Journal of Computational and Graphical Statistics*, vol. 17, no. 2, pp. 426–443, 2008.
- [14] M. Abraham and F. Wu, "Empirical likelihood on the full parameter space," *The Annals of Statistics*, vol. 41, no. 4, pp. 2176–2196, 2013.
- [15] W. Liang, H. Dai, and S. He, "Mean empirical likelihood," *Computational Statistics & Data Analysis*, vol. 138, pp. 155–169, 2019.
- [16] N. Reisby, L. F. Gram, P. Bech et al., "Imipramine: clinical effects and pharmacokinetic variability," *Psychopharmacology*, vol. 54, no. 3, pp. 263–272, 1977.