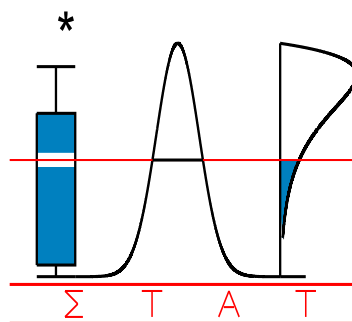# T E C H N I C A L
R E P O R T

0441

EMPIRICAL LIKELIHOOD IN A SEMIPARAMETRIC MODEL FOR MISSING RESPONSE DATA

WANG, L. and N. VERAVERBEKE

I A P   S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

Empirical likelihood in a semi-parametric model for missing response data [*]

Lichun Wang,[†] Noel Veraverbeke

Center for Statistics, Limburgs Universitair Centrum
Universitaire Campus, B-3590 Diepenbeek, Belgium

Abstract Let Y be a response and, given covariate X, Y has a conditional density $f(y|x, \theta)$, where θ is a unknown p-dimensional vector of parameters and the marginal distribution of X is unknown. When responses are missing at random, with auxiliary information and imputation, we define an adjusted empirical log-likelihood ratio for the mean of Y and obtain its asymptotic distribution. A simulation study is conducted to compare the adjusted empirical log-likelihood and the normal approximation method in terms of coverage accuracies.

1. Introduction

Consider a semi-parametric model which parameterizes the conditional density $f(y|x, \theta)$ of Y given X, where θ is a p-dimensional vector of parameters and the marginal distribution of X is unknown. We are interested in inference on the mean of the response variable Y, say μ, but due to all kinds of reasons such as loss of information, or failure on obtaining all data, the responses are often missing. In fact, incomplete data can appear in many situations such as market research surveys, mail enquiries and so on. In practice, one often obtains a random sample of incomplete data as

$$(X_i, Y_i, \delta_i), \quad i = 1, \cdots, n, \tag{1}$$

where all d-dimensional vectors X_i are observed and $\delta_i = 0$ if Y_i is missing, otherwise $\delta_i = 1$. At such occasion, one usual method is to drop X_i with the missing response from our analysis, but obviously this can cause a serious loss of efficiency when a substantial proportion of X_i is lost. Another technique is to impute a value for the missing Y_i so as to obtain complete data and proceed the statistical process. There are many imputation methods for missing responses such as kernel regression imputation used by Cheng (1994) and Wang and Rao (2002a), linear regression imputation adopted by Wang and Rao (2001) and ratio imputation which appeared in Rao (1996), and so forth.

In this paper, we impute the missing response by a maximum likelihood estimate (MLE). Using the incomplete data (X_i, Y_i, δ_i), $1 \le i \le n$, and an auxiliary information, we adopt the empirical likelihood technique, introduced by Owen (1988, 1990), to make inference on the mean of Y. It is well known that the empirical likelihood is a nonparametric method which is very useful for constructing confidence regions or intervals for the mean and other parameters. It has many advantages over some modern and classical methods such as the bootstrap method and the normal-approximation-based method. One advantage of using empirical likelihood is that the shape of the confidence region is determined automatically by the data. Also, there is an excellent exposition of the empirical likelihood in Hall and La Scala (1990). Some related work can be found in DiCiccio and Romano (1989), Qin and Lawless (1994), Chen (1993, 1994) and Wang and Rao (2001, 2002a, 2002b), among others.

This paper is organized as follows. In Section 2, we define the adjusted empirical log-likelihood ratio for the mean of the response and introduce the main results. Some simulation results will be exhibited in Section 3 to compare the empirical likelihood method with the normal-approximation-based method. For convenience,

we convey the proofs of the main results to Section 4, which is at the end of the paper.

2. Main results

Throughout this paper, we assume that Y is missing at random (MAR), that is, $P(\delta = 1|Y, X) = P(\delta = 1|X)$. Furthermore, assume that there is an auxiliary information of the form $E\phi(X, \theta) = 0$, where $\phi(X, \theta)$ is a $q \times 1$ known vector-valued function of X and θ and $q > p$.

Let $m(x, \theta) = E(Y|X = x)$. Note that $Em(X_i, \theta) = EY_i$, so we can impute Y_i by $m(X_i, \theta)$ and estimate μ by

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}[\delta_i Y_i + (1 - \delta_i)m(X_i, \theta)] \hat{=} \frac{1}{n}\sum_{i=1}^{n}\hat{Y}_i \tag{2}$$

when Y_i is missing. Hence, under MAR, $E\hat{Y}_i = \mu$ if μ is the true parameter. Thus, the problem of testing whether μ is the true parameter is equivalent to testing whether $E\hat{Y}_i = \mu$ for $i = 1, \cdots, n$. This motivates us to define the empirical log-likelihood ratio (Owen (1990)) as

$$l(\mu) = -2\max\sum_{i=1}^{n}\log(n\omega_i), \tag{3}$$

where the maximum is taken over all sets of nonnegative numbers $\omega_1, \cdots, \omega_n$ satisfying $\sum_{i=1}^{n}\omega_i = 1$ and $\sum_{i=1}^{n}\omega_i\hat{Y}_i = \mu$. Clearly, $l(\mu)$ contains not only μ but also the unknown parameter θ, hence, it cannot be applied directly to make inference on μ. To solve this problem, we first need to estimate θ.

Based on the auxiliary information and the observed data (X_i, Y_i, δ_i), $1 \le i \le n$, we maximize the likelihood

$$L(\theta) = \prod_{i=1}^{n}[f(Y_i|X_i, \theta)dG(X_i)]^{\delta_i}[dG(X_i)]^{1-\delta_i} = \prod_{i=1}^{n}p_i\prod_{i=1}^{n}f^{\delta_i}(Y_i|X_i, \theta) \tag{4}$$

3

subject to $\sum_{i=1}^{n} p_i = 1$ and $\sum_{i=1}^{n} p_i \phi(X_i, \theta) = 0$, where $p_i = dG(X_i)$ and $G(x)$ denotes the unknown marginal distribution of X.

If 0 is in the convex hull of $\phi(X_1, \theta), \cdots, \phi(X_n, \theta)$, by the Lagrange multiplier method, we can easily get

$$p_i = \frac{1}{n}[1 + \lambda_1^T \phi(X_i, \theta)]^{-1}, \quad i = 1, \cdots, n,$$

where λ_1 is the Lagrange multiplier. It satisfies the following equation

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\phi(X_i, \theta)}{1 + \lambda_1^T \phi(X_i, \theta)} = 0. \tag{5}$$

Substituting p_i into (4), we have

$$\log L(\theta) = -\sum_{i=1}^{n} \log[n(1 + \lambda_1^T \phi(X_i, \theta))] + \sum_{i=1}^{n} \delta_i \log f(Y_i | X_i, \theta) \tag{6}$$

with λ_1 being the solution of the equation (5).

Assume that the combined maximum likelihood estimate (MLE) $\bar{\theta}_n$ satisfies

$$\frac{\partial \log L(\theta)}{\partial \theta} = -\sum_{i=1}^{n} \frac{1}{1 + \lambda_1^T \phi(X_i, \theta)} \frac{\partial \phi(X_i, \theta)}{\partial \theta^T} \lambda_1 + \sum_{i=1}^{n} \delta_i \frac{\partial \log f(Y_i | X_i, \theta)}{\partial \theta} = 0, \tag{7}$$

where we use the fact that

$$\sum_{i=1}^{n} \frac{1}{1 + \lambda_1^T \phi(X_i, \theta)} \frac{\partial \lambda_1}{\partial \theta^T} \phi(X_i, \theta) = 0.$$

Then, we define an estimated empirical log-likelihood ratio by

$$\bar{l}_n(\mu) = -2 \max_{\sum_{i=1}^{n} \omega_i = 1, \sum_{i=1}^{n} \omega_i Y_{in} = \mu} \sum_{i=1}^{n} \log(n\omega_i), \tag{8}$$

where $Y_{in} = \delta_i Y_i + (1 - \delta_i) m(X_i, \bar{\theta}_n)$. Using Lagrange multipliers, when $\min_{1 \leq i \leq n} Y_{in} < \mu < \max_{1 \leq i \leq n} Y_{in}$, we have

$$\omega_i = \frac{1}{n}[1 + \lambda_2(Y_{in} - \mu)]^{-1}, \quad i = 1, \cdots, n,$$

4

where λ_2 is the solution of

$$\frac{1}{n}\sum_{i=1}^{n}\frac{Y_{in}-\mu}{1+\lambda_2(Y_{in}-\mu)}=0. \tag{9}$$

Hence, we have

$$\bar{l}_n(\mu)=2\sum_{i=1}^{n}\log[1+\lambda_2(Y_{in}-\mu)]. \tag{10}$$

Since the Y_{in} in (10) are not independent and identically distributed (i.i.d), $\bar{l}_n(\mu)$ is asymptotically a non-standard chi-square variable. In fact, it can be shown that $\bar{l}_n(\mu)$, multiplied by some population quantity, follows a chi-quare distribution with one degree of freedom. In other words, $r(\mu)\bar{l}_n(\mu) \sim \chi_1^2$ asymptotically. Hence, in order to use this result for constructing a confidence interval for the mean μ, one has to estimate the coefficient $r(\mu)$.

Define an adjusted empirical log-likelihood ratio by

$$\hat{l}_{n,ad}(\mu)=\hat{r}(\mu)\bar{l}_n(\mu), \tag{11}$$

where

$$\hat{r}(\mu)=\frac{\hat{S}_2(\mu)}{\hat{S}_1(\mu)},$$

with

$$\hat{S}_2(\mu)=\frac{1}{n}\sum_{i=1}^{n}(Y_{in}-\mu)^2, \tag{12}$$

and

$$\hat{S}_1(\mu)=\hat{S}_{11}(\mu)+\hat{S}_{12}(\mu)+\hat{S}_{13}(\mu), \tag{13}$$

where

$$\hat{S}_{11}(\mu) = \frac{1}{n}\sum_{i=1}^{n}\{\delta_i[Y_i-m(X_i,\bar{\theta}_n)]\}^2+\frac{1}{n}\sum_{i=1}^{n}[m(X_i,\bar{\theta}_n)-\mu]^2$$

$$+\left[\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)m^{(1)}(X_i,\bar{\theta}_n)\right]^T\hat{\Gamma}^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}(1+\delta_i)m^{(1)}(X_i,\bar{\theta}_n)\right],$$

$$\hat{S}_{12}(\mu) = \left[\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)m^{(1)}(X_i,\bar{\theta}_n)\right]^T \hat{\Gamma}^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial\phi(X_i,\theta)}{\partial\theta^T}|_{\theta=\bar{\theta}_n}\right]$$

$$\times\left[\frac{1}{n}\sum_{i=1}^{n}\phi(X_i,\bar{\theta}_n)\phi^T(X_i,\bar{\theta}_n)\right]^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial\phi(X_i,\theta)}{\partial\theta}|_{\theta=\bar{\theta}_n}\right]\hat{\Gamma}^{-1}$$

$$\times\left[\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)m^{(1)}(X_i,\bar{\theta}_n)\right],$$

$$\hat{S}_{13}(\mu) = -2\left[\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)m^{(1)}(X_i,\bar{\theta}_n)\right]^T \hat{\Gamma}^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial\phi(X_i,\theta)}{\partial\theta^T}|_{\theta=\bar{\theta}_n}\right]$$

$$\times\left[\frac{1}{n}\sum_{i=1}^{n}\phi(X_i,\bar{\theta}_n)\phi^T(X_i,\bar{\theta}_n)\right]^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}\phi(X_i,\bar{\theta}_n)m(X_i,\bar{\theta}_n)\right]. \qquad (14)$$

where

$$\hat{\Gamma} = \left\{\frac{1}{n}\sum_{i=1}^{n}\delta_i\frac{\partial\log f(Y_i|X_i,\theta)}{\partial\theta}\frac{\partial\log f(Y_i|X_i,\theta)}{\partial\theta^T}\right\}_{\theta=\bar{\theta}_n}$$

and $m^{(1)}(x,\theta)$ denotes the first order partial derivative with respect to θ.

In what follows, we shall establish a theorem for the adjusted empirical log-likelihood defined in (11), which is a nonparametric version of Wilks's theorem. Before stating the theorem, we first make the following assumptions.

(A1) $f(y|x,\theta)$ satisfies the regularity conditions on the asymptotic normality of the MLE in fully parametric models.

(A2) $\phi(x,\theta)$ satisfies the regularity conditions as $\psi(x,\theta)$ specified in Qin and Lawless (1994).

(A3) The first order derivative with respect to θ of the left side of $\int yf(y|x,\theta)dy = m(x,\theta)$ can be obtained by differentiating under the integral sign.

(A4) $EY^2 < \infty$.

(A5) $E||\phi(X,\theta)||^2 < \infty$.

(A6) $E||m^{(1)}(X,\theta)||^2 < \infty$.

(A7) Each element of the matrix $m^{(2)}(X,\theta)$ has a finite second order moment, where $m^{(2)}(x,\theta)$ denotes the second order partial derivative with respect to θ.

(A8) Each element of the matrix $\partial^2 \log f(Y|X,\theta)/\partial\theta\partial\theta^T$ has a finite second order moment.

Now, we have the following main result. Its proof will be given in Section 4.

Theorem 1. *Under the assumptions (A1)-(A8), if μ is the true parameter, we have*

$$\hat{l}_{n,ad}(\mu) \xrightarrow{\mathcal{L}} \chi_1^2.$$

Hence, a simple approach to construct an approximate $1 - \alpha$ confidence interval for the mean μ, based on Theorem 1, is

$$I_\alpha = \{\mu : \hat{l}_{n,ad}(\mu) \leq c_\alpha\} \tag{15}$$

with $P(\chi_1^2 \leq c_\alpha) = 1 - \alpha$. Clearly, I_α will have the correct coverage probability $1 - \alpha$ asymptotically, i.e.

$$P(\mu \in I_\alpha) = 1 - \alpha + o(1). \tag{16}$$

On the other hand, from (2)-(7), we can define an empirical likelihood-based estimator of μ as

$$\bar{\mu} = \frac{1}{n}\sum_{i=1}^n \frac{\delta_i Y_i + (1-\delta_i)m(X_i,\bar{\theta}_n)}{1 + \lambda_1^T\phi(X_i,\bar{\theta}_n)} = \frac{1}{n}\sum_{i=1}^n \frac{Y_{in}}{1 + \lambda_1^T\phi(X_i,\bar{\theta}_n)}. \tag{17}$$

Then, we have the following theorem.

Theorem 2. *Under the assumptions (A1)-(A8), if μ is the true parameter, we have*

$$\sqrt{n}(\bar{\mu} - \mu) \xrightarrow{\mathcal{L}} N(0, S_{au}(\mu)),$$

where $S_{au}(\mu) = S_{11}(\mu) + S_{12}(\mu) - E[m(X,\theta)\phi^T(X,\theta)][E\phi(X,\theta)\phi^T(X,\theta)]^{-1}E[m(X,\theta)\phi(X,\theta)]$, with

$$
\begin{aligned}
S_{11}(\mu) &= \{E[(1 - P(X))m^{(1)}(X,\theta)]\}^T\Gamma^{-1}\{E[(1 + P(X))m^{(1)}(X,\theta)]\} \\
&\quad + E[P(X)\,Var(Y|X)] + Var(m(X,\theta)) \tag{18}
\end{aligned}
$$

7

and

$$S_{12}(\mu) = \{E[(1-P(X))m^{(1)}(X,\theta)]\}^T \Gamma^{-1} E[\frac{\partial \phi(X,\theta)}{\partial \theta^T}][E\phi(X,\theta)\phi^T(X,\theta)]^{-1}$$

$$\times E[\frac{\partial \phi(X,\theta)}{\partial \theta}]\Gamma^{-1} E[(1-P(X))m^{(1)}(X,\theta)], \qquad (19)$$

where $P(X) = P(\delta = 1|X)$, and

$$\Gamma = E\left[P(X)E\left(\frac{\partial \log f(Y|X,\theta)}{\partial \theta}\frac{\partial \log f(Y|X,\theta)}{\partial \theta^T}|X\right)\right]. \qquad (20)$$

Let

$$\hat{S}_{au}(\mu) = -\left[\frac{1}{n}\sum_{i=1}^{n}m(X_i,\bar{\theta}_n)\phi^T(X_i,\bar{\theta}_n)\right]\left[\frac{1}{n}\sum_{i=1}^{n}\phi(X_i,\bar{\theta}_n)\phi^T(X_i,\bar{\theta}_n)\right]^{-1}$$

$$\times \left[\frac{1}{n}\sum_{i=1}^{n}m(X_i,\bar{\theta}_n)\phi(X_i,\bar{\theta}_n)\right] + \hat{S}_{11}(\mu) + \hat{S}_{12}(\mu),$$

where $\hat{S}_{11}(\mu)$, $\hat{S}_{12}(\mu)$ are defined in (14). Then, we can give the following normal approximation-based $1 - \alpha$ confidence interval for μ: $\bar{\mu} \pm u_{\alpha/2}\hat{S}_{au}^{1/2}(\mu)/\sqrt{n}$, where $u_{\alpha/2}$ is the upper $\alpha/2$ percentile point of the standard normal distribution.

3. Simulations

In this Section, we shall use the empirical likelihood method and the normal-approximation method, which are based on Theorem 1 and Theorem 2 respectively, to construct confidence intervals for the mean μ.

For the model

$$f(y|x,\theta) = (2\pi)^{-1/2}\exp[-(y-\theta_1-\theta_2 x)^2/2], \quad (\theta_1,\theta_2) = (1.0, 0.5), \quad X \sim N(1,1),$$

we adopt the following auxiliary information,

$$\phi(X,\theta) = \mu - \theta_1 - \theta_2 X.$$

8

Consider the following two response probability functions under the MAR assumption.

Case 1: $P(\delta = 1 | X = x) = 0.6$ for all x;

Case 2: $P(\delta = 1 | X = x) = 0.8 + 0.2|x - 1|$ if $|x - 1| \leq 1$, and 0.95 elsewhere.

It is easy to see that $Y \sim N(\mu, 1.25)$ and $m(x, \theta) = \theta_1 + \theta_2 x$. To compare the empirical likelihood with the normal-approximation method, first we know that the assumptions (A1)-(A8) are satisfied with the above model. Then, we need to give the combined MLE $\bar{\theta}_n$ for the parameter θ, and calculate $\hat{S}_1(\mu)$, $\hat{S}_2(\mu)$ and $\hat{S}_{au}(\mu)$, respectively. At last, combining the proof of the Theorem 1 with (11), we compute $\hat{l}_{n,ad}(\mu)$.

First, as a contrast, when there is no auxiliary information, we generate, respectively, 5000 Monte Carlo random samples of size n=20,40, 60, 100 for Case 1 and Case 2. Based on the coverage probability for μ, the results are presented in Table 1 and Table 2.

Table 1 Case 1—Coverage probabilities for μ

n	Nominal level is 0.90		Nominal level is 0.95	
	normal approximation	Empirical likelihood	normal approximation	Empirical likelihood
20	0.9276	0.9136	0.9726	0.9642
40	0.8918	0.8948	0.9454	0.9506
60	0.8746	0.8914	0.9350	0.9402
100	0.8668	0.8854	0.9256	0.9364

Table 2 Case 2—Coverage probabilities for μ

n	Nominal level is 0.90		Nominal level is 0.95	
	normal approximation	Empirical likelihood	normal approximation	Empirical likelihood
20	0.9070	0.9102	0.9642	0.9652
40	0.8995	0.9051	0.9500	0.9556
60	0.8908	0.9042	0.9444	0.9508
100	0.8926	0.9000	0.9430	0.9436

In Table 1, due to $E[P(\delta = 1|X = x)] = 0.6$, when n is small(i.e. n=20), we find that the normal-approximation-based method is a bit better than the empirical likelihood, but in Table 2, when $E[P(\delta = 1|X = x)](\approx 0.90)$ is increased, obviously, the normal-approximation-based method is inferior to the empirical likelihood method for all sample sizes.

When there is auxiliary information, we generate, respectively, 5000 Monte Carlo random samples for size n=20, 50, 100 in Case 1, and, respectively, 5000 Monte Carlo random samples for size n=5, 10, 20, 50, in Case 2. Based on the coverage probability for μ, the results are presented in Table 3 and Table 4.

From Table 3, we find that the empirical likelihood is better than the normal approximation method for all sample sizes. Note that the Theorem 2 is also based on imputation, therefore, it is obvious that the auxiliary information can cause a substantial gain of efficiency of coverage accuracy. Also, note that $E[P(\delta = 1|X = x)] = 0.6$ is relatively small in Case 1. In the case that $E[P(\delta = 1|X = x)](\approx 0.90)$ is increased, by Table 4, we see that empirical likelihood method is still overmatching the normal approximation method more or less. Furthemore, from Table 3 and Table 4, it is also interesting to note that the coverage accuracies for both of these two methods generally tend to increase as the sample size n gets larger. However, this not always not the case. This reason is that $(x_i, y_i)'s$ are different for each different sample size n as well as MLE $\bar{\theta}_n$ and hence this makes the comparisons under different sizes more difficult.

Table 3 Case 1—Coverage probabilities for μ

	Nominal level is 0.90		Nominal level is 0.95	
n	normal approximation	Empirical likelihood	normal approximation	Empirical likelihood
20	0.9004	0.9642	0.9574	0.9864
50	0.9172	0.9598	0.9620	0.9882
100	0.9228	0.9658	0.9668	0.9916

Table 4   Case 2—Coverage probabilities for μ

n	Nominal level is 0.90		Nominal level is 0.95	
	normal approximation	Empirical likelihood	normal approximation	Empirical likelihood
5	0.8040	0.8922	0.8534	0.9364
10	0.8792	0.9132	0.9204	0.9620
20	0.9036	0.9188	0.9526	0.9676
50	0.9240	0.9184	0.9598	0.9614

4.   Proofs of Theorems

We need the following lemmas.

Lemma 1. *Under the assumptions (A1)-(A3), (A5) and (A7)-(A8), if μ is the true parameter, we have*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Y_{in} - \mu) \xrightarrow{\mathcal{L}} N(0, S_1(\mu)),$$

where $S_1(\mu) = S_{11}(\mu) + S_{12}(\mu) + S_{13}(\mu)$ with

$$
\begin{aligned}
S_{13}(\mu) &= -2\{E[(1 - P(X))m^{(1)}(x,\theta)]\}^T \Gamma^{-1} E[\frac{\partial \phi(X,\theta)}{\partial \theta^T}] \\
&\times E[\phi(X,\theta)\phi^T(X,\theta)]^{-1} E[\phi(X,\theta)m(X,\theta)].
\end{aligned}
\tag{21}
$$

Proof. Write

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Y_{in} - \mu) = \sqrt{n}(Q_{1n} + Q_{2n} + Q_{3n}), \tag{22}$$

where

$$
\begin{aligned}
Q_{1n} &= \frac{1}{n}\sum_{i=1}^{n}\delta_i[Y_i - m(X_i,\theta)], \\
Q_{2n} &= \frac{1}{n}\sum_{i=1}^{n}[m(X_i,\theta) - \mu], \\
Q_{3n} &= \frac{1}{n}\sum_{i=1}^{n}(1 - \delta_i)[m(X_i,\bar{\theta}_n) - m(X_i,\theta)].
\end{aligned}
$$

11

Since Q_{1n} and Q_{2n} are means of i.i.d. random variables, the main task is to consider Q_{3n}. First, using (A5) and mimicking the proof of the theorem 1 in Owen (1990), we have

$$||\lambda_1|| = O_p(n^{-1/2}), \tag{23}$$

where $||.||$ denotes the Euclidean norm. Note that $\bar{\theta}_n$ satisfies the equation (7), so that by (A7) and Taylor expansion, one can show that Q_{3n} is asymptotically equivalent to

$$- \frac{1}{n} \{E[(1-\delta)m^{(1)}(X,\theta)]\}^T A^{-1} [\sum_{i=1}^{n} \delta_i \frac{\partial \log f(Y_i|X_i,\theta)}{\partial \theta}$$
$$- \sum_{i=1}^{n} \frac{1}{1 + \lambda_1^T \phi(X_i,\theta)} \frac{\partial \phi(X_i,\theta)}{\partial \theta^T} \lambda_1], \tag{24}$$

where

$$A = E \left[\delta \frac{\partial^2 \log f(Y|X,\theta)}{\partial\theta\partial\theta^T} \right].$$

Second, expanding (5)

$$0 = \frac{1}{n} \sum_{i=1}^{n} \frac{\phi(X_i,\theta)}{1 + \lambda_1^T \phi(X_i,\theta)} = \frac{1}{n} \sum_{i=1}^{n} \phi(X_i,\theta) - \frac{1}{n} \sum_{i=1}^{n} \phi(X_i,\theta)\phi^T(X_i,\theta)\lambda_1$$
$$+ \frac{1}{n} \sum_{i=1}^{n} \frac{\phi(X_i,\theta)(\lambda_1^T \phi(X_i,\theta))^2}{1 + \lambda_1^T \phi(X_i,\theta)}, \tag{25}$$

with the final term bounded by

$$\left| \frac{1}{n} \sum_{i=1}^{n} \frac{\phi(X_i,\theta)(\lambda_1^T \phi(X_i,\theta))^2}{1 + \lambda_1^T \phi(X_i,\theta)} \right| = O_p(n^{-1})o_p(n^{\frac{1}{2}})O_p(1) = o_p(n^{-\frac{1}{2}}),$$

yields

$$\lambda_1 = \left[\sum_{i=1}^{n} \phi(X_i,\theta)\phi^T(X_i,\theta) \right]^{-1} \sum_{i=1}^{n} \phi(X_i,\theta) + o_p(n^{-\frac{1}{2}}). \tag{26}$$

Finally, by assumptions (A8) and (A3), we have, respectively

$$A = -\Gamma \tag{27}$$

and

$$E\left[P(X)E\left(\frac{\partial \log f(Y|X,\theta)}{\partial \theta}Y|X\right)\right] = E[P(X)m^{(1)}(X,\theta)]. \qquad (28)$$

Lemma 1 follows from (22), (24) and (26)-(28).

Lemma 2. *Under the assumptions (A1)-(A2), (A4) and (A6)-(A7), if μ is the true parameter, we have*

$$\hat{S}_2(\mu) = \frac{1}{n}\sum_{i=1}^{n}(Y_{in} - \mu)^2 \xrightarrow{p} S_2(\mu),$$

where

$$S_2(\mu) = E[p(X)\,Var(Y|X)] + Var(m(X,\theta)). \qquad (29)$$

Proof. Similar to (22), we have

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}(Y_{in} - \mu)^2 &= \frac{1}{n}\sum_{i=1}^{n}\delta_i^2[Y_i - m(X_i,\theta)]^2 + \frac{1}{n}\sum_{i=1}^{n}[m(X_i,\theta) - \mu]^2 \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)^2[m(X_i,\bar{\theta}_n) - m(X_i,\theta)]^2 \\
&\quad + \frac{2}{n}\sum_{i=1}^{n}\delta_i[Y_i - m(X_i,\theta)][m(X_i,\theta) - \mu] \\
&\quad + \frac{2}{n}\sum_{i=1}^{n}\delta_i(1-\delta_i)[Y_i - m(X_i,\theta)][m(X_i,\bar{\theta}_n) - m(X_i,\theta)] \\
&\quad + \frac{2}{n}\sum_{i=1}^{n}(1-\delta_i)[m(X_i,\theta) - \mu][m(X_i,\bar{\theta}_n) - m(X_i,\theta)] \\
&\hat{=}\ R_{1n} + R_{2n} + R_{3n} + R_{4n} + R_{5n} + R_{6n}. \qquad (30)
\end{aligned}$$

It is easy to see that

$$R_{1n} \xrightarrow{p} E[p(X)\mathrm{Var}(Y|X)] \qquad (31)$$

and

$$R_{2n} \xrightarrow{p} \mathrm{Var}(m(X,\theta)). \qquad (32)$$

13

Following a proof analogous to (24) and by the assumptions (A6) and (A7), we have

$$R_{3n} \xrightarrow{p} 0. \tag{33}$$

Obviously,

$$R_{4n} \xrightarrow{p} 0. \tag{34}$$

Since $E[\delta(1-\delta)(Y - m(X,\theta))m^{(k)}(X,\theta)] = 0$ $(k = 1, 2)$, one has

$$R_{5n} \xrightarrow{p} 0. \tag{35}$$

Also, by the assumptions (A4), (A6) and (A7), we can obtain

$$E[(1 - P(X))(m(X,\theta))^i m^{(k)}(X,\theta)] < \infty, \quad \text{for} \quad i = 0, 1, \quad k = 1, 2.$$

Hence

$$R_{6n} \xrightarrow{p} 0. \tag{36}$$

Then, from (30)-(36), the conclusion of Lemma 2 is held.

Lemma 3. *Under the assumptions (A1)-(A2), (A4) and (A6), we have*

$$\max_{1 \leq i \leq n} |Y_{in}| = o_p(n^{\frac{1}{2}}).$$

Proof. Notice

$$\max_{1 \leq i \leq n} |Y_{in}| \leq \max_{1 \leq i \leq n} |Y_i| + \max_{1 \leq i \leq n} |m(X_i, \theta)| + \max_{1 \leq i \leq n} |m(X_i, \bar{\theta}_n) - m(X_i, \theta)|. \tag{37}$$

By $E|m(X,\theta)|^2 \leq EY^2 < \infty$ and Owen (1990)'s Lemma 3, we obtain

$$\max_{1 \leq i \leq n} |Y_i| = o_p(n^{\frac{1}{2}}), \quad \max_{1 \leq i \leq n} |m(X_i, \theta)| = o_p(n^{\frac{1}{2}}). \tag{38}$$

Furthermore, by (A1)-(A2) and $E||m^{(1)}(X,\theta)||^2 < \infty$, we have

$$\max_{1 \leq i \leq n} |m(X_i, \bar{\theta}_n) - m(X_i, \theta)| \leq \max_{1 \leq i \leq n} ||m^{(1)}(X_i, \theta^*)|| ||\bar{\theta}_n - \theta|| = o_p(n^{\frac{1}{2}}), \tag{39}$$

where $||\theta^* - \theta|| \leq ||\bar{\theta}_n - \theta||$. This together with (38) proves Lemma 3.

Lemma 4. *If μ is the true parameter and $E||m^{(1)}(X, \theta)||^2 < \infty$, then*

$$\lim_{n \to \infty} P(\min_{1 \leq i \leq n} Y_{in} < \mu < \max_{1 \leq i \leq n} Y_{in}) = 1.$$

Proof. Since the proof is similar to Wang and Rao (2002a), we omit it.

Lemma 5. *Under the conditions of Lemma 1 and Lemma 2, we have*

$$\lambda_2 = O_p(n^{-1/2}).$$

Proof. By Lemma 1, we have

$$\frac{1}{n} \sum_{i=1}^{n} Y_{in} = O_p(n^{-1/2}).$$

This together with the same arguments used in the proof of (23) proves Lemma 5.

Proof of Theorem 1. From (10), by Taylor expansion, we have

$$\bar{l}_n(\mu) = 2 \sum_{i=1}^{n} [\lambda_2 (Y_{in} - \mu)] - \sum_{i=1}^{n} [\lambda_2 (Y_{in} - \mu)]^2 + 2 \sum_{i=1}^{n} \xi_i, \qquad (40)$$

where, by Lemma 2, Lemma 3 and Lemma 5, one has $|2 \sum_{i=1}^{n} \xi_i| = o_p(1)$.

Moreover, expanding (9)

$$0 = \frac{1}{n} \sum_{i=1}^{n} \frac{Y_{in} - \mu}{1 + \lambda_2 (Y_{in} - \mu)} = \frac{1}{n} \sum_{i=1}^{n} (Y_{in} - \mu) - \frac{\lambda_2}{n} \sum_{i=1}^{n} (Y_{in} - \mu)^2$$

$$+ \frac{\lambda_2^2}{n} \sum_{i=1}^{n} \frac{(Y_{in} - \mu)^3}{1 + \lambda_2 (Y_{in} - \mu)}, \qquad (41)$$

where

$$\left| \frac{\lambda_2^2}{n} \sum_{i=1}^{n} \frac{(Y_{in} - \mu)^3}{1 + \lambda_2 (Y_{in} - \mu)} \right| = O_p(n^{-1}) o_p(n^{\frac{1}{2}}) O_p(1) = o_p(n^{-\frac{1}{2}}),$$

we obtain

$$\lambda_2 \sum_{i=1}^{n} (Y_{in} - \mu) = \sum_{i=1}^{n} [\lambda_2 (Y_{in} - \mu)]^2 + o_p(1) \qquad (42)$$

15

and

$$\lambda_2 = \left[\sum_{i=1}^{n}(Y_{in} - \mu)^2\right]^{-1}\sum_{i=1}^{n}(Y_{in} - \mu) + o_p(n^{-\frac{1}{2}}). \tag{43}$$

Combining (40), (42) and (43), we have

$$\bar{l}_n(\mu) = \lambda_2^2\sum_{i=1}^{n}(Y_{in} - \mu)^2 + o_p(1) = \left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Y_{in} - \mu)\right]^2\left[\frac{1}{n}\sum_{i=1}^{n}(Y_{in} - \mu)^2\right]^{-1} + o_p(1). \tag{44}$$

Recalling the definition of $\hat{l}_{n,ad}(\mu)$, we get

$$\hat{l}_{n,ad}(\mu) = \left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{Y_{in} - \mu}{\sqrt{\hat{S}_1(\mu)}}\right]^2 + o_p(1). \tag{45}$$

Then, by a similar proof of Lemma 2, we can obtain

$$\hat{S}_1(\mu) \xrightarrow{p} S_1(\mu). \tag{46}$$

where $S_1(\mu)$ is defined in Lemma 1.

Hence, Theorem 1 follows from (45), (46) and Lemma 1.

Proof of Theorem 2. From (17), write

$$\sqrt{n}(\bar{\mu} - \mu) = \sqrt{n}(I_{1n} + I_{2n}), \tag{47}$$

where

$$I_{1n} = \frac{1}{n}\sum_{i=1}^{n}\frac{Y_{in}}{1 + \lambda_1^T\phi(X_i, \theta)} - \mu, \tag{48}$$

$$I_{2n} = \frac{1}{n}\sum_{i=1}^{n}\frac{Y_{in}}{1 + \lambda_1^T\phi(X_i, \bar{\theta}_n)} - \frac{1}{n}\sum_{i=1}^{n}\frac{Y_{in}}{1 + \lambda_1^T\phi(X_i, \theta)}. \tag{49}$$

First, using Lemma 3 and (A5), we have

$$I_{1n} = \frac{1}{n}\sum_{i=1}^{n}(Y_{in} - \mu) - \frac{1}{n}\sum_{i=1}^{n}Y_{in}\phi^T(X_i, \theta)\lambda_1 + o_p(n^{-1/2}). \tag{50}$$

Then, by (26), we get

$$
\begin{aligned}
I_{1n} &= \frac{1}{n}\sum_{i=1}^{n}(Y_{in}-\mu) - \frac{1}{n}\sum_{i=1}^{n}Y_{in}\phi^T(X_i,\theta)\left[\frac{1}{n}\sum_{i=1}^{n}\phi(X_i,\theta)\phi^T(X_i,\theta)\right]^{-1} \\
&\quad \times \frac{1}{n}\sum_{i=1}^{n}\phi(X_i,\theta) + o_p(n^{-1/2}).
\end{aligned}
\tag{51}
$$

In order to obtain the asymptotic normality of $\sqrt{n}I_{1n}$, we first need prove

$$
\frac{1}{n}\sum_{i=1}^{n}Y_{in}\phi^T(X_i,\theta) \xrightarrow{p} E[m(X,\theta)\phi^T(X,\theta)].
\tag{52}
$$

Write

$$
\frac{1}{n}\sum_{i=1}^{n}Y_{in}\phi^T(X_i,\theta) = J_{1n} + J_{2n} + J_{3n},
\tag{53}
$$

where

$$
J_{1n} = \frac{1}{n}\sum_{i=1}^{n}\delta_i[Y_i - m(X_i,\theta)]\phi^T(X_i,\theta),
\tag{54}
$$

$$
J_{2n} = \frac{1}{n}\sum_{i=1}^{n}m(X_i,\theta)\phi^T(X_i,\theta),
\tag{55}
$$

$$
J_{3n} = \frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)[m(X_i,\bar{\theta}_n) - m(X_i,\theta)]\phi^T(X_i,\theta).
\tag{56}
$$

It is readily seen that

$$
J_{1n} \xrightarrow{p} 0,
\tag{57}
$$

and

$$
J_{2n} \xrightarrow{p} E[m(X,\theta)\phi^T(X,\theta)].
\tag{58}
$$

Moreover, by the assumptions (A5)-(A7) and the same arguments used in the proof of (24), we get

$$
J_{3n} \xrightarrow{p} 0.
\tag{59}
$$

Hence, (52) holds.

Then, together with (51) and (52), we have

$$
\sqrt{n}I_{1n} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Y_{in} - \mu) - E[m(X,\theta)\phi^T(X,\theta)][E\phi(X,\theta)\phi^T(X,\theta)]^{-1}
$$

$$
\times \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi(X,\theta) + o_p(1). \tag{60}
$$

It follows Lemma 1

$$
Cov\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Y_{in} - \mu), E[m(X,\theta)\phi^T(X,\theta)][E\phi(X,\theta)\phi^T(X,\theta)]^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi(X,\theta)\right)
$$

$$
= \{E[(1 - P(X))m^{(1)}(X,\theta)]\}^T A^{-1} E[\frac{\partial\phi(X,\theta)}{\partial\theta^T}]
$$

$$
\times [E\phi(X,\theta)\phi^T(X,\theta)]^{-1} E[m(X,\theta)\phi(X,\theta)]
$$

$$
+ E[m(X,\theta)\phi^T(X,\theta)][E\phi(X,\theta)\phi^T(X,\theta)]^{-1} E[m(X,\theta)\phi(X,\theta)]. \tag{61}
$$

Hence, by Lemma 1 and $A = -\Gamma$, we know

$$
\sqrt{n}I_{1n} \xrightarrow{\mathcal{L}} N(0, S_{au}(\mu)). \tag{62}
$$

On the other hand, from Lemma 3 and (23), by the Theorem 1 stated in Qin (2000), we have

$$
|\sqrt{n}I_{2n}| = \left|\frac{1}{n}\sum_{i=1}^{n}Y_{in}\frac{\lambda_1^T[\phi(X_i,\theta) - \phi(X_i,\bar{\theta}_n)]}{[1 + \lambda_1^T\phi(X_i,\bar{\theta}_n)][1 + \lambda_1^T\phi(X_i,\theta)]}\right|
$$

$$
\leq \frac{1}{n}\sum_{i=1}^{n}|Y_{in}|\frac{|\lambda_1^T\phi^{(1)}(X_i,\theta^*)\sqrt{n}(\theta - \bar{\theta}_n)|}{[1 + \lambda_1^T\phi(X_i,\bar{\theta}_n)][1 + \lambda_1^T\phi(X_i,\theta)]}
$$

$$
= o_p(n^{1/2})O_p(n^{-1/2})O_p(1)
$$

$$
= o_p(1), \tag{63}
$$

where $\phi^{(1)}(x,\theta)$ denotes the first order partial derivative with respect to θ and $||\theta^* - \theta|| \leq ||\bar{\theta}_n - \theta||$.

Hence, the conclusion of Theorem 2 follows from (47), (62) and (63).

18

References

Chen, S. X., 1993. On the accuracy of empirical likelihood confidence regions for linear regression model. Ann. Inst. Statist. Math. 45, 621-637.

Chen, S. X., 1994. Empirical likelihood confidence intervals for linear regression coefficients. J. Multivariate Anal. 49, 24-40.

Cheng, P. E., 1994. Nonparametric estimation of mean functionals with data missing as random. J. Amer. Statist. Assoc. 89, 81-87.

DiCiccio, T. J., Romano, J. P., 1989. On adjustments based on the signed root of the empirical likelihood ratio statistic. Biometrika 76, 447-456.

Hall, P., La Scala, B., 1990. Methodology and algorithms of empirical likelihood. Int. Statist. Rev. 58, 109-127.

Owen, A., 1988. Empirical likelihood ratio confidence intervals for a single functional. Biometrika 75, 237-249.

Owen, A., 1990. Empirical likelihood ratio confidence regions. Ann. Statist. 18, 90-120.

Qin, J., Lawless, J. F., 1994. Empirical likelihood and general estimating equations. Ann. Statist. 22, 300-325.

Qin, J., 2000. Combining parametric and empirical likelihoods. Biometrika 87, 484-490.

Rao, J. N. K., 1996. On variance estimation with imputed survey data. J. Amer. Statist. Assoc. 91, 499-502.

Wang, Q. H., Rao, J. N. K., 2001. Empirical likelihood for linear regres sion models under imputation for missing responses. The Canadian Journal of Statistics 29, 597-608.

Wang, Q. H., Rao, J. N. K., 2002a. Empirical likelihood-based inference under imputation for missing response data. Ann. Statist. 30, 896-924.

Wang, Q. H., Rao, J. N. K., 2002b. Empirical likelihood-based inference in linear errors-in-covariables models with validation data. Biometrika 89, 345-358.