Uniform improvement of empirical likelihood for missing response problem

Kwun Chuen Gary Chan

Department of Biostatistics and Department of Health Services
University of Washington
Seattle WA, USA
e-mail: kcgchan@uw.edu

Abstract: An empirical likelihood (EL) estimator was proposed by Qin and Zhang (2007) for improving the inverse probability weighting estimation in a missing response problem. The authors showed by simulation studies that the finite sample performance of EL estimator is better than certain existing estimators and they also showed large sample results for the estimator. However, the empirical likelihood estimator does not have a uniformly smaller asymptotic variance than other existing estimators in general. We consider several modifications to the empirical likelihood estimator and show that the proposed estimator dominates the empirical likelihood estimator and several other existing estimators in terms of asymptotic efficiencies under missing at random. The proposed estimator also attains the minimum asymptotic variance among estimators having influence functions in a certain class and enjoys certain double robustness properties.

Keywords and phrases: Auxiliary information, double robustness, empirical likelihood, missing data, survey calibration.

Received February 2011.

1. Introduction and existing estimators

Suppose we are interested in estimating the mean μ of a random variable Y but Y is partially observed subject to missingness. Let X be a vector of covariates that are fully observable and R be an indicator that Y is observed. The observed data are $(r_i, r_i y_i, x_i)$ for $i = 1, \ldots, n$ and are *i.i.d.* realizations from (R, RY, X). Under a missing at random assumption that $P(R = 1|Y, X) = P(R = 1|X) = \pi_0(X)$, μ can be consistently estimated by the inverse probability weighting (IPW) estimator

$$\hat{\mu}_{IPW} = \frac{1}{n} \sum_{i=1}^{n} \frac{r_i}{\pi_0(x_i)} y_i$$

For missing data applications the non-missing probability is usually not known but is being modeled. Suppose $P(R = 1|X) = \pi(X; \beta_0)$, where β_0 is a finite dimensional parameter. Based on $(r_1, x_1), \ldots, (r_n, x_n)$, the parameter β_0 can be estimated by solving a likelihood score equation $n^{-1} \sum_{i=1}^{n} \dot{l}(x_i; \beta) = 0$ where $\dot{l}(x; \beta) = [1 - \pi(x; \beta)]^{-1} [r_i - \pi(x; \beta)] \frac{\partial \pi}{\partial \beta}(x; \beta)$, and we denote $\hat{\beta}$ to be the solution. We usually replace $\pi_0(x_i)$ by the estimated probability $\pi(x_i; \hat{\beta})$ in IPW estimation.

The IPW estimator is intuitive and easy to implement but is inefficient in general, because information from X is not fully utilized when Y is not observed. To improve efficiency, an empirical likelihood estimator is proposed by Qin and Zhang (2007) defined as

$$\hat{\mu}_{EL} = \sum_{i=1}^{n} r_i p_i^{EL} y_i$$

where weights p_i^{EL} are defined for complete case observations (*i.e.* when $r_i = 1$) and the following empirical log-likelihood function

$$l = \sum_{i=1}^{n} r_i \log p_i^{EL}$$

is maximized subject to constraints

$$p_i^{EL} \geq 0$$
$$\sum_{i=1}^{n} r_i p_i^{EL} = 1$$
$$\sum_{i=1}^{n} r_i p_i^{EL} \pi(x_i; \hat{\beta}) = \hat{\theta}$$
$$\sum_{i=1}^{n} r_i p_i^{EL} a(x_i; \hat{\beta}) = \hat{a}$$

where $a = (a_1, \ldots, a_p)$ is a fixed vector function of $p < n$ dimensions, $\hat{\theta} = n^{-1} \times \sum_{i=1}^{n} \pi(x_i; \hat{\beta})$ and $\hat{a} = n^{-1} \times \sum_{i=1}^{n} a(x_i)$. Let $s(x; \beta, \theta, a) = \{1 - \theta\pi^{-1}(x; \beta), \pi^{-1}(x; \beta)[a(x) - a]^T\}^T$ and $n_1 = \sum_{i=1}^{n} r_i$. Solving the constrained maximization problem, the empirical likelihood weights p_i^{EL} are expressed in terms of a vector of Lagrange multipliers $\hat{\lambda}_{EL}$

$$p_i^{EL} = \frac{1}{n_1} \frac{\hat{\theta}\pi^{-1}(x_i; \beta)}{1 + \hat{\lambda}_{EL}^T s(x_i; \hat{\beta}, \hat{\theta}, \hat{a})} \tag{1.1}$$

and the Lagrange multipliers satisfies a system of estimating equations

$$\sum_{i=1}^{n} \frac{r_i s(x_i; \hat{\beta}, \hat{\theta}, \hat{a})}{1 + \hat{\lambda}_{EL}^T s(x_i; \hat{\beta}, \hat{\theta}, \hat{a})} = 0. \tag{1.2}$$

Information from incomplete observations are utilized implicitly in the construction of weights p_i^{EL} from the constraints. When Y and $a(X)$ are correlated, the empirical likelihood estimator usually improves upon the IPW estimator in terms of estimation efficiency. Although the EL methods works for arbitrary choice of $a(X)$, Qin and Zhang (2007) showed that optimal efficiency is attained when the conditional expectation $E(Y|X)$ is a linear combination of $a(X)$. In

practice, one could model $E(Y|X)$ from the observed data, but the optimal case cannot be achieved because no model is perfect. We therefore consider the general case in which $E(Y|X)$ may not be a linear combination of $a(X)$ and a broader optimality result is shown under the general case. The main contribution of the paper is to propose a uniform improvement of the existing empirical likelihood estimator for arbitrary pre-specified $a(X)$.

We note that the empirical likelihood estimator of Qin and Zhang (2007) is different from the general empirical likelihood methodologies of Qin and Lawless (1994) for estimation from overidentified system of estimating equations and optimally combining estimating equations. Unlike the single-step estimator discussed in Qin and Lawless (1994), the empirical likelihood estimator of Qin and Zhang (2007) is computed in a two-step manner which would be much easier to implement. However, two-step estimators may have different efficiency compare to single-step estimators. Following Qin and Lawless (1994), Qin, Zhang and Leung (2009) studied a single-step empirical likelihood estimator for missing data problems. Other empirical likelihood estimators has also been proposed by Wang and Rao (2002) and Wang and Chen (2009), among others, for missing data problems under different settings. In missing data analysis, there is often an interest in studying robust estimator that gives consistent estimates under certain conditions even when the missing data mechanism is misspecified, a so-called doubly robustness property. The empirical likelihood estimator of Qin and Zhang (2007) enjoys the double robustness property together with the proposed modified estimator but double robustness was not discussed in Qin, Zhang and Leung (2009). We will further discuss this point in section 5. For the purpose of clarity, we refer empirical likelihood estimator to the estimator of Qin and Zhang (2007) in the rest of the paper.

The empirical likelihood estimator has nice small sample properties shown in simulations, but it does not theoretically dominate other existing estimators in terms of asymptotic efficiency for arbitrary pre-specified $a(X)$. A partial review of other existing methods can be found in Kang and Schafer (2007). To study asymptotic efficiencies of estimators, we consider estimators which influence functions belong to the following class

$$\mathcal{L} = \left\{ \frac{R}{\pi_0(X)}[Y - m(X)] + [m(X) - \mu] : m(X) \text{ is linear in } \tilde{a}(X) \right\}.$$

where $\tilde{a} = (1, a_1, \ldots, a_p, (1 - \pi_0)^{-1}\partial\pi^T/\partial\beta)^T$. The class $\mathcal{L}$ is a large class of influence functions and contains many important existing estimators. The EL estimator, together with the survey calibration estimator [Deville and Särndal (1992)] and the augmented inverse probability weighting estimator [Robins, Rotnitzky and Zhao (1994)], all have influence functions in the class $\mathcal{L}$. However, all of the aforementioned estimators do not attain the minimum variance in class $\mathcal{L}$ and do not strictly dominates one another under missing at random assumptions and for estimated weights. The main purpose of this paper is to study modifications of the empirical likelihood estimator that attains the minimum asymptotic variance among class $\mathcal{L}$, and therefore theoretically dominate

the EL estimator and many other existing estimators when the same amount of covariate information is used. To motivate our main results, section 2 will first discuss the optimality of empirical likelihood under the stronger missing completely at random assumption. In section 3, we will first show that the EL estimator is suboptimal in general under missing at random with estimated missing probability. We go on to discuss modifications that improve efficiencies and achieve optimality conditions. Section 4 will present simulation studies comparing the finite performance of estimators. The choice of class $\mathcal{L}$ among other possible classes of influence functions are further discussed in section 5.

2. Optimality of empirical likelihood under missing completely at random

To better motivate the modifications of the EL estimator under missing at random assumption, we first investigate the optimality of EL estimator under the stronger assumption of missing completely at random, *i.e.* $P(R = 1|Y, X) = \pi_0$, a constant independent of Y and X. We assume that π_0 is a known constant in this section. Under this special case, the EL weight is

$$p_i^{EL} = \frac{1}{n_1} \frac{1}{1 + \hat{\lambda}_{EL}^T s'(x; \hat{a})} \tag{2.1}$$

where $s'(x; a) = \{\pi_0^{-1}[a(x) - a]^T\}^T$ and the Lagrange multipliers satisfy the following system of estimating equations

$$\sum_{i=1}^{n} \frac{r_i s'(x_i; \hat{a})}{1 + \hat{\lambda}_{EL}^T s(x_i; \hat{a})} = 0. \tag{2.2}$$

Under this special case, we consider estimators which influence functions belong to the following class

$$\mathcal{L}' = \left\{ \frac{R}{\pi_0}[Y - m(X)] + [m(X) - \mu] : m(X) \text{ is linear in } a'(X) \right\}$$

where $a'(X) = (1, a(X)^T)^T$.

Before we show that empirical likelihood attains the minimum efficiency among class $\mathcal{L}'$, we first characterize the conditions for $m(X)$ to achieve minimum efficiency. For an estimator having influence function in class $\mathcal{L}'$, its asymptotic variance is

$$
\begin{aligned}
& n^{-1} \times Var\left\{ \frac{R}{\pi_0}[Y - m(X)] + [m(X) - \mu] \right\} \\
=\ & n^{-1} \times Var\left(E\left\{ \frac{R}{\pi_0}[Y - m(X)] + [m(X) - \mu]|Y, X \right\} \right) \\
& + n^{-1} \times E\left(Var\left\{ \frac{R}{\pi_0}[Y - m(X)] + [m(X) - \mu]|Y, X \right\} \right) \\
=\ & n^{-1} \times \left\{ Var(Y) + \frac{1 - \pi_0}{\pi_0} E[Y - m(X)]^2 \right\}
\end{aligned}
$$

Since $m(X) = c^T a'(X)$ in class $\mathcal{L}'$, the asymptotic variance of estimators in class $\mathcal{L}'$ is minimized at

$$n^{-1} \times \left(Var(Y) + \frac{1 - \pi_0}{\pi_0} \min_{c \in \mathcal{R}^{p+1}} E\{[Y - c^T a'(X)]^2\} \right). \tag{2.3}$$

To characterize the optimal estimator, let $m_0(X) = c_0^T a'(X)$ where

$$c_0 = \arg \min_{c \in \mathcal{R}^{p+1}} E\{[Y - c^T a'(X)]^2\}$$

By the definition of c_0, the following set of orthogonal conditions are satisfied:

$$E(a'(X)[Y - c_0^T a'(X)]) = E(a'(X)[Y - m_0(X)]) = 0 \tag{2.4}$$

and the asymptotic variance of $\frac{R}{\pi_0}[Y - m_0(X)] + [m_0(X) - \mu]$ is minimized among the class $\mathcal{L}'$. Furthermore, we show the optimality of empirical likelihood estimator by noting that

$$
\begin{aligned}
\hat{\mu}_{EL} - \mu &= \sum_{i=1}^{n} r_i p_i^{EL} y_i - \mu \\
&= \sum_{i=1}^{n} r_i p_i^{EL} (y_i - m_0(x_i)) + \sum_{i=1}^{n} r_i p_i^{EL} m_0(x_i) - \mu \\
&= \sum_{i=1}^{n} r_i p_i^{EL} (y_i - m_0(x_i)) + \frac{1}{n} \sum_{i=1}^{n} (m_0(x_i) - \mu) \\
&= \frac{1}{n} \sum_{i=1}^{n} r_i \left[\frac{n/n_1}{(1 + \hat{\lambda}_{EL}^T s'(x_i; \hat{a}))} - \frac{1}{\pi_0} \right] (y_i - m_0(x_i)) \\
&\quad + \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{r_i}{\pi_0} [y_i - m_0(x_i)] + [m_0(x_i) - \mu] \right\} \tag{2.5}
\end{aligned}
$$

where the second last equality follows from (2.1) and (2.2). Note that the second summation in (2.5) corresponds to the optimal estimator in class $\mathcal{L}'$. To show that the empirical likelihood estimator is optimal within class $\mathcal{L}'$, we need to show that the first summation in (2.5) is $o_p(n^{-1/2})$. Under mild regularity conditions, we can show by asymptotic properties for estimating equations [Newey and McFadden (1994)] that $\sqrt{n}(\hat{\lambda}_{EL} - 0)$ converges weakly to a Gaussian distribution. By Taylor series expansion,

$$\frac{1}{n} \sum_{i=1}^{n} \frac{r_i}{\pi_0} \left(\frac{1}{1 + \hat{\lambda}_{EL}^T s'(x_i; \hat{a})} - 1 \right) (y_i - m_0(x_i))$$

$$= A_{EL,MCAR}(\hat{\lambda}_{EL} - 0) + \frac{1 - \pi_0}{\pi_0} \lambda^T|_{\lambda=0}(\hat{a}^T - \mu_a^T) + d \left(\frac{n}{n_1} - \frac{1}{\pi_0} \right) + o_p(n^{-1/2})$$

where $\mu_a = E(a(X))$, $A_{EL,MCAR} = E((a'(X) - (1, \mu_a^T)^T)(Y - m_0(X)))$ and $d = E(Y - m_0(X))$. Note that $A_{EL,MCAR} = 0$ and $d = 0$ follow from the

orthogonal condition (2.4). Therefore we can conclude that

$$\hat{\mu}_{EL} - \mu = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{r_i}{\pi_0}[y_i - m_0(x_i)] + [m_0(x_i) - \mu]\right\} + o_p(n^{-1/2})$$

and hence $\hat{\mu}_{EL}$ is asymptotically equivalent to the minimum variance estimator in class $\mathcal{L}'$.

3. Modified empirical likelihood under missing at random with estimated missing probability

In section 2 we showed that empirical likelihood is optimal in which it attains the minimum asymptotic variance in a class $\mathcal{L}'$ under the missing completely at random assumption. However, empirical likelihood estimator does not attain the minimum asymptotic variance in class $\mathcal{L}$ under the missing at random assumption. We can see this fact by considering similar arguments as in section 2. First, an estimator having influence function in class $\mathcal{L}$ has asymptotic variance given by

$$n^{-1} \times Var\left\{\frac{R}{\pi_0(X)}[Y - m(X)] + [m(X) - \mu]\right\}$$

$$= n^{-1} \times Var\left(E\left\{\frac{R}{\pi_0(X)}[Y - m(X)] + [m(X) - \mu]|Y,X\right\}\right)$$

$$+ n^{-1} \times E\left(Var\left\{\frac{R}{\pi_0(X)}[Y - m(X)] + [m(X) - \mu]|Y,X\right\}\right)$$

$$= n^{-1} \times \left\{Var(Y) + E\left[\frac{1 - \pi_0(X)}{\pi_0(X)}(Y - m(X))^2\right]\right\}$$

Since $m(X) = c^T\tilde{a}(X)$ in class $\mathcal{L}$, the minimum asymptotic variance in class $\mathcal{L}$ is

$$n^{-1} \times \left\{Var(Y) + \min_{c\in\mathbb{R}^q} E\left[\frac{1 - \pi_0(X)}{\pi_0(X)}(Y - c^T\tilde{a}(X))^2\right]\right\} \qquad (3.1)$$

where q is the dimension of $\tilde{a}$.

Next, we characterize the minimum variance estimator in class $\mathcal{L}$. Let $m_0(X) = c_0^T\tilde{a}(X)$ where

$$c_0 = \arg\min_{c\in\mathbb{R}^q} E\left(\frac{1 - \pi_0(X)}{\pi_0(X)}(Y - c^T\tilde{a}(X))^2\right)$$

where q is the dimension of $\tilde{a}$. The variance of $\frac{R}{\pi_0(X)}[Y - m_0(X)] + [m_0(X) - \mu]$ is the minimum among the class $\mathcal{L}$ as in (3.1). Note that the optimal estimator will have $\{1 - \pi_0(X)\}/\pi_0(X)$ inside the expectation (3.1) whereas $(1 - \pi_0)/\pi_0$ can be written outside the expectation in (2.3). By the definition of c_0, the following set of orthogonal conditions are satisfied.

$$E\left(\frac{1 - \pi_0(X)}{\pi_0(X)}\tilde{a}^T(X)(Y - c_0^T\tilde{a}(X))\right) = E\left(\frac{1 - \pi_0(X)}{\pi_0(X)}\tilde{a}^T(X)(Y - m_0(X))\right) = 0.$$

$$(3.2)$$

We then show that, unlike under the special case of missing completely at random, the empirical likelihood estimator do not achieve minimum variance among class $\mathcal{L}$ under missing at random. We note that

$$
\begin{aligned}
\hat{\mu}_{EL} - \mu &= \sum_{i=1}^{n} r_i p_i^{EL} y_i - \mu \\
&= \sum_{i=1}^{n} r_i p_i^{EL}(y_i - m_0(x_i)) + \sum_{i=1}^{n} r_i p_i^{EL} m_0(x_i) - \mu \\
&= \sum_{i=1}^{n} r_i p_i^{EL}(y_i - m_0(x_i)) + \frac{1}{n} \sum_{i=1}^{n}(m_0(x_i) - \mu) \\
&= \frac{1}{n} \sum_{i=1}^{n} r_i \left[\frac{n\hat{\theta}/n_1}{\pi(x_i; \hat{\beta})(1 + \hat{\lambda}_{EL}^T s(x_i; \hat{\beta}, \hat{a}, \hat{b}))} - \frac{1}{\pi_0(x_i)} \right] (y_i - m_0(x_i)) \\
&\quad + \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{r_i}{\pi_0(x_i)}[y_i - m_0(x_i)] + [m_0(x_i) - \mu] \right\}
\end{aligned}
\tag{3.3}
$$

As shown in Qin and Zhang (2007), $\sqrt{n}(\hat{\lambda}^{EL} - 0)$ converges weakly to Gaussian distributions. By Taylor Series expansions,

$$
\begin{aligned}
&\frac{1}{n} \sum_{i=1}^{n} r_i \left[\frac{n\hat{\theta}/n_1}{\pi(x_i; \hat{\beta})(1 + \hat{\lambda}_{EL}^T s(x_i; \hat{\beta}, \hat{a}))} - \frac{1}{\pi_0(x_i)} \right] (y_i - m_0(x_i)) \\
&= A_{EL}(\hat{\lambda}^{EL} - 0) + B(\hat{\beta} - \beta_0) + c\lambda^T|_{\lambda=0}(\hat{a}^T - \mu_a^T)^T + d(\hat{\theta} - \theta_0) + o_p(n^{-1/2})
\end{aligned}
\tag{3.4}
$$

where $A_{EL} = -E(\pi_0^{-1}(X)(a''(X) - \mu_{a''})^T(Y - m_0(X)))$, $a''(X) = \{\pi(X), a(X)^T\}^T$, $\mu_{a''} = E(a''(X))$, $B = -E(\pi_0^{-1}(X)\frac{\partial \pi^T}{\partial \beta}(Y - m_0(X)))$, $c = E(\frac{1 - \pi_0(X)}{\pi_0(X)})$, $\mu_a = E(a(X))$ and $d = E(\pi_0(X))^{-1}E(Y - m_0(X))$. Unlike the case of missing completely at random, there is no direct relationship between A_{EL} and the orthogonal conditions (3.2), and the matrix is generally nonzero.

Note that even when the missing probability is known, A_{EL} is non-zero in general and therefore the empirical likelihood estimator is still suboptimal among class $\mathcal{L}$.

Our aim is to construct an estimator attaining the minimum asymptotic variance (3.1). When the missing probability is estimated, we employ the following three strategies in constructing the optimal estimator: (1) We need to modify both the empirical likelihood weights and the estimating equations for finding Lagrange multipliers so that the corresponding matrix A_{MEL} defined later is a zero matrix. (2) We need to augment our number of constraints to ensure that $\sum_{i=1}^{n} r_i p_i^{MEL} m_0(x_i) = \frac{1}{n} \sum_{i=1}^{n} m_0(x_i)$. (3) We want to remove the influence from the estimate $\hat{\theta}$ by removing it entirely in the estimation. To achieve these goals, we consider the following modifications. First, we start from point (3). Since $\hat{\theta}$ is a consistent estimator for $P(R = 1)$, in the modification we replace

$\hat{\theta}/n_1$ in (1.1) by $1/n$. Next, we replace $s(x_i; \hat{\beta}, \hat{\theta}, \hat{a})$ by an augmented version $s^*(x_i; \hat{\beta}, \hat{a}, \hat{b})$, where

$$s^*(x; \beta, a, b) = \frac{1 - \pi(x; \beta)}{\pi(x; \beta)} \times \left(1, [a(x) - a]^T, \left[\frac{1}{1 - \pi(x; \beta)} \frac{\partial \pi}{\partial \beta}(x; \beta) - b\right]^T\right)^T$$

and $\hat{b} = \sum_{i=1}^n (1 - \pi(x_i; \hat{\beta}))^{-1} \partial \pi(x_i; \hat{\beta}) / \partial \beta$. The modified empirical likelihood estimator is defined as

$$\hat{\mu}_{MEL} = \sum_{i=1}^n r_i p_i^{MEL} y_i$$

where the modified empirical likelihood (MEL) weights are

$$p_i^{MEL} = \frac{1}{n} \frac{\pi^{-1}(x_i; \hat{\beta})}{1 + \hat{\lambda}_{MEL}^T s^*(x_i; \hat{\beta}, \hat{a}, \hat{b})} \tag{3.5}$$

and the pseudo Lagrange multiplier $\hat{\lambda}_{MEL}$ are obtained by solving

$$\sum_{i=1}^n \frac{r_i(1 - \pi(x_i; \hat{\beta}))^{-1} s^*(x; \hat{\beta}, \hat{a}, \hat{b})}{1 + \hat{\lambda}_{MEL}^T s^*(x_i; \hat{\beta}, \hat{a}, \hat{b})} = e \tag{3.6}$$

where $e = (1, 0, \ldots, 0)^T$. Plugging (3.5) into (3.6) gives $\sum_{i=1}^n r_i p_i^{MEL} = 1$ and $\sum_{i=1}^n r_i p_i^{MEL} a(x_i) = \hat{a}$, which corresponds to the constraints in empirical likelihood estimation. In addition, we have $\sum_{i=1}^n r_i p_i^{MEL} (1 - \pi(x_i; \hat{\beta}))^{-1} \partial \pi(x_i; \hat{\beta}) / \partial \beta = \hat{b}$. Unlike (1.1) and (1.2), (3.5) and (3.6) are not implied by constrained maximization problems. The reason is similar to the fact that not all estimating functions are derivative of log-likelihood functions.

We now show that the construction of modified empirical likelihood estimator mentioned above attains minimum asymptotic variance among class $\mathcal{L}$. Again, we note that

$$\begin{aligned}
\hat{\mu}_{MEL} - \mu &= \sum_{i=1}^n r_i p_i^{MEL} y_i - \mu \\
&= \sum_{i=1}^n r_i p_i^{MEL} (y_i - m_0(x_i)) + \sum_{i=1}^n r_i p_i^{MEL} m_0(x_i) - \mu \\
&= \sum_{i=1}^n r_i p_i^{MEL} (y_i - m_0(x_i)) + \frac{1}{n} \sum_{i=1}^n (m_0(x_i) - \mu) \\
&= \frac{1}{n} \sum_{i=1}^n r_i \left[\frac{1}{\pi(x_i; \hat{\beta})(1 + \hat{\lambda}_{MEL}^T s^*(x_i; \hat{\beta}, \hat{a}, \hat{b}))} - \frac{1}{\pi_0(x_i)}\right] (y_i - m_0(x_i)) \\
&\quad + \frac{1}{n} \sum_{i=1}^n \left\{\frac{r_i}{\pi_0(x_i)} [y_i - m_0(x_i)] + [m_0(x_i) - \mu]\right\}
\end{aligned} \tag{3.7}$$

where the second last equality follows from (3.5), (3.6) and the definition of s^*. Under mild regularity conditions [Newey and McFadden (1994)], it can be shown that $\sqrt{n}(\hat{\lambda}^{MEL} - 0)$ and $\sqrt{n}(\hat{\beta} - \beta)$ converges weakly to Gaussian distributions. Also by Taylor Series expansions,

$$\frac{1}{n}\sum_{i=1}^{n} r_i \left[\frac{1}{\pi(x_i;\hat{\beta})(1+\hat{\lambda}_{MEL}^T s^*(x_i;\hat{\beta},\hat{a},\hat{b}))} - \frac{1}{\pi_0(x_i)} \right] (y_i - m_0(x_i))$$
$$= A_{MEL}(\hat{\lambda}_{MEL} - 0) + B(\hat{\beta} - \beta) + c\lambda^T|_{\lambda=0}(\hat{a}^T - \mu_a^T, \hat{b}^T - \mu_b^T)^T + o_p(n^{-1/2})$$
(3.8)

where $A_{MEL} = -E(\frac{1-\pi_0(X)}{\pi_0(X)}(\tilde{a}(X) - \mu_{\tilde{a}})^T(Y - m_0(X)))$, $\mu_{\tilde{a}} = E(\tilde{a}(X))$, $B = -E(\frac{1}{\pi_0(X)}\frac{\partial\pi^T}{\partial\beta}(Y - m_0(X)))$, $c = E(\frac{1-\pi_0(X)}{\pi_0(X)})$, $\mu_a = E(a(X))$ and $\mu_b = E((1 - \pi_0(X))^{-1}\frac{\partial\pi(X;\beta_0)}{\partial\beta})$. For modified empirical likelihood, matrices A_{MEL} and B are both 0 following the orthogonal conditions (3.2). Therefore, it follows from (3.7) and (3.8) that the influence function of $\hat{\mu}_{MEL}$ is $\frac{R}{\pi_0(X)}[Y - m_0(X)] + [m_0(X) - \mu]$ which attains the minimum variance among the class $\mathcal{L}$.

For the special case where $E(Y|X) = b_0 + b_1^T a(X)$ for some b_0 and b_1, $E[(Y - E(Y|X))^2|X]$ is minimized at each X and therefore $m_0(X) = E(Y|X)$. In this case, the modified empirical likelihood estimator attains the semiparametric efficiency bound. Also, empirical likelihood attains the same semiparametric efficiency bound under correct specification of the outcome regression model. However, for arbitrary pre-specified $a(X)$, the modified empirical likelihood has a smaller asymptotic variance than empirical likelihood and other estimators having influence functions in class $\mathcal{L}$ in general.

The modified empirical likelihood estimator also possesses a double robustness property as for the empirical likelihood estimator. Suppose $E(Y|X) = b_0 + b_1^T a(X) = m_0(X)$ but the missing data model $\pi(x;\beta)$ is misspecified. The estimates $\hat{\beta}$, $\hat{\lambda}_{MEL}$ and $\hat{b}$ converges in probability to some constants β^*, λ^* [White (1982), Hall and Inoue (2003)] and μ_b^* and λ^* is usually non-zero. From (3.7) we note that

$$\hat{\mu}_{MEL} = \frac{1}{n}\sum_{i=1}^{n} \frac{r_i}{\pi(x_i;\hat{\beta})(1+\hat{\lambda}_{MEL}^T s^*(x_i;\hat{\beta},\hat{a},\hat{\beta}))}(y_i - m_0(x_i)) + \frac{1}{N}\sum_{i=1}^{N} m_0(x_i)$$
$$\xrightarrow{p} E\left(\frac{\pi_0(X)}{\pi(X;\beta^*)(1+\lambda^{*T}s^*(X;\beta^*,\mu_a,\mu_b^*))}(E(Y|X) - m_0(X))\right)$$
$$+ E(E(Y|X))$$
$$= 0 + \mu = \mu$$

That is, the modified empirical likelihood estimator is consistent when the outcome regression model is correctly specified even when the missing data model is misspecified.

4. Simulations

In this section we present simulation studies to evaluate the finite sample performance of the modified empirical likelihood estimator, together with the EL estimator and two other existing estimators: a survey calibration (CAL) estimator and an augmented inverse probability weighting (AIPW) estimator. The empirical likelihood calibration estimator of Chen, Sitter and Wu (2002) is defined as

$$\hat{\mu}_{CAL} = \sum_{i=1}^{n} r_i p_i^{CAL} y_i$$

where the calibration weights p_i^{CAL} are expressed in terms of a vector of Lagrange multipliers $\hat{\lambda}_{CAL}$

$$p_i^{CAL} = \frac{\pi^{-1}(x_i; \hat{\beta})[\sum_{i=1}^{n} r_i \pi^{-1}(x_i; \hat{\beta})]^{-1}}{1 + \hat{\lambda}_{CAL}^T [a(x_i) - \hat{a}]}$$

and the Lagrange multipliers satisfies

$$\sum_{i=1}^{n} \frac{r_i \pi^{-1}(x_i; \hat{\beta})[a(x_i) - \hat{a}]}{1 + \hat{\lambda}_{CAL}^T [a(x_i) - a]} = 0.$$

The calibration estimator and the EL estimator is very similar but not identical and they have different asymptotic variance in general.

We also compare the proposed estimator with the augmented inverse probability weighting (AIPW) estimator proposed in the literature to improve efficiency of IPW estimation [Robins, Rotnitzky and Zhao (1994)]. To implement the AIPW estimator, a regression model is fitted using the complete case subsample, treating Y as outcome and $a(X)$ as covariates. We let $\hat{m}(X)$ to be the prediction from the fitted model. An AIPW estimator is defined as

$$\hat{\mu}_{AIPW} = \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{r_i}{\pi(x_i; \hat{\beta})} y_i + \left[\frac{r_i - \pi(x_i; \hat{\beta})}{\pi(x_i; \hat{\beta})} \right] \hat{m}(x_i) \right\}$$

The first simulation study followed the scenario in Kang and Schafer (2007) for estimating a population mean. The scenario was carefully designed so that the assumed outcome regression and missing data models are nearly correct under misspecification, but the AIPW estimator can be severely biased. Sample sizes for each simulated data set was 200, 500 or 1000, and 1000 Monte Carlo datasets were generated. For each observation, a random vector $Z = (Z_1, Z_2, Z_3, Z_4)$ was generated from a standard multivariate normal distribution, and transformations $X_1 = \exp(Z_1/2), X_2 = Z_2/(1 + \exp(Z_1)), X_3 = (Z_1 Z_3/25 + 0.6)^3$ and $X_4 = (Z_2 + Z_4 + 20)^2$ were defined. The outcome of interest Y was generated from a normal distribution with mean $210 + 27.4Z_1 + 13.7Z_2 + 13.7Z_3 + 13.7Z_4$ and unit variance, and Y was observed with probability $\exp(\eta_0(Z))/(1 + \exp(\eta_0(Z)))$ where $\eta_0(Z) = -Z_1 + 0.5Z_2 - 0.25Z_3 - 0.1Z_4$. The

Comparisons among estimators under the Kang and Schafer scenario with four possible combinations of correct and misspecified missing data and outcome regression models, (a) both correct, (b) correct missing data model and incorrect outcome regression, (c) incorrect missing data model but correct outcome regression and (d) both incorrect. RMSE represents the square root of sampling mean squared error. RE represents the relative efficiency compared to the modified empirical likelihood estimator

n		n=200			n=500			n=1000		
		Bias	RMSE	RE	Bias	RMSE	RE	Bias	RMSE	RE
(a)	$\hat{\mu}_{AIPW}$	0.02	2.50	1.00	0.03	1.62	1.02	0.01	1.13	1.00
	$\hat{\mu}_{EL}$	0.02	2.50	1.00	0.03	1.62	1.01	0.01	1.13	1.00
	$\hat{\mu}_{CAL}$	0.02	2.50	1.00	0.03	1.62	1.01	0.01	1.13	1.00
	$\hat{\mu}_{MEL}$	0.02	2.50	1.00	0.03	1.61	1.00	0.01	1.13	1.00
(b)	$\hat{\mu}_{AIPW}$	0.28	3.77	2.07	0.12	2.74	2.76	0.06	1.65	2.02
	$\hat{\mu}_{EL}$	0.49	2.90	1.23	0.30	1.78	1.16	0.19	1.22	1.11
	$\hat{\mu}_{CAL}$	0.28	3.14	1.44	0.16	1.94	1.38	0.10	1.35	1.35
	$\hat{\mu}_{MEL}$	0.21	2.62	1.00	0.16	1.65	1.00	0.11	1.16	1.00
(c)	$\hat{\mu}_{AIPW}$	0.01	2.55	1.04	0.11	2.36	2.12	−0.01	1.25	1.22
	$\hat{\mu}_{EL}$	0.02	2.50	.100	0.03	1.62	1.00	0.01	1.13	1.00
	$\hat{\mu}_{CAL}$	0.01	2.49	0.99	0.03	1.61	0.99	0.01	1.13	1.00
	$\hat{\mu}_{MEL}$	0.03	2.50	1.00	0.03	1.62	1.00	0.01	1.13	1.00
(d)	$\hat{\mu}_{AIPW}$	−8.00	41.07	>100	−39.66	898.58	>100	−13.38	73.39	>100
	$\hat{\mu}_{EL}$	−1.71	3.52	1.01	−2.06	2.81	1.72	−2.15	2.52	2.15
	$\hat{\mu}_{CAL}$	−2.73	4.82	1.89	−3.53	4.62	4.66	−4.16	5.05	8.62
	$\hat{\mu}_{MEL}$	−1.07	3.51	1.00	−1.06	2.14	1.00	−1.18	1.72	1.00

correctly specified outcome and missing data models were regression models with Z as covariates, whereas we treated X to be the covariates in misspecified models instead of Z. Kang and Schafer (2007) showed that the missspecified models are nearly correctly specified. In each case we considered four possible combinations of correct and misspecified missing data and outcome regression models: (a) both correct; (b) correct missing data model and incorrect outcome regression; (c) incorrect missing data model but correct outcome regression and (d) both incorrect. For correctly specified outcome model, $a(Z) = (Z_1, Z_2, Z_3, Z_4)$ and $a(X) = (X_1, X_2, X_3, X_4)$ for misspecified outcome model. We compared the performances of the augmented inverse probability weighted estimator $\hat{\mu}_{AIPW}$, the empirical likelihood estimator $\hat{\mu}_{EL}$, the survey calibration estimator $\hat{\mu}_{CAL}$ and the modified empirical likelihood estimator $\hat{\mu}_{MEL}$. Relative efficiency (RE) is defined as the ratio of mean squared error of an estimator to the mean squared error of the modified empirical likelihood estimator. The results are shown in Table 1.

Simulation results showed that EL, CAL, AIPW and MEL estimators all had relatively small bias when either the missing data model or the outcome regression model was correctly specified. When both models were correctly specified, all estimators had very similar performances because all of them were semiparametric locally efficient. When only one of the two models were correctly specified, the empirical likelihood, calibration and modified empirical likelihood estimators were more efficient than the AIPW estimator. When both models were misspecified, the AIPW estimator had a considerable bias and variability but the other empirical likelihood based estimators showed much better per-

Comparisons among estimators under the Qin and Zhang scenario with four possible
combinations of correct and misspecified missing data and outcome regression models, (a)
both correct, (b) correct missing data model and incorrect outcome regression, (c) incorrect
missing data model but correct outcome regression and (d) both incorrect. RMSE represents
the square root of sampling mean squared error. RE represents the relative efficiency
compared to the modified empirical likelihood estimator

		n=200			n=500			n=1000		
		Bias	RMSE	RE	Bias	RMSE	RE	Bias	RMSE	RE
(a)	$\hat{\mu}_{AIPW}$	−0.01	0.31	1.00	<0.01	0.20	1.08	<0.01	0.14	1.02
	$\hat{\mu}_{EL}$	−0.01	0.31	1.00	<0.01	0.19	1.00	<0.01	0.14	1.00
	$\hat{\mu}_{CAL}$	−0.01	0.31	1.00	<0.01	0.19	1.00	<0.01	0.14	1.00
	$\hat{\mu}_{MEL}$	−0.02	0.31	1.00	<0.01	0.19	1.00	<0.01	0.14	1.00
(b)	$\hat{\mu}_{AIPW}$	<0.01	0.39	1.53	0.01	0.25	1.76	<0.01	0.17	1.61
	$\hat{\mu}_{EL}$	−0.02	0.37	1.41	0.01	0.23	1.43	−0.01	0.17	1.53
	$\hat{\mu}_{CAL}$	−0.01	0.38	1.49	0.01	0.23	1.50	−0.01	0.17	1.59
	$\hat{\mu}_{MEL}$	−0.01	0.31	1.00	<0.01	0.19	1.00	<0.01	0.14	1.00
(c)	$\hat{\mu}_{AIPW}$	−0.01	0.31	1.00	<0.01	0.20	1.08	<0.01	0.14	1.02
	$\hat{\mu}_{EL}$	−0.01	0.31	.100	<0.01	0.19	1.00	<0.01	0.14	1.00
	$\hat{\mu}_{CAL}$	−0.01	0.31	1.00	<0.01	0.19	1.00	<0.01	0.14	1.00
	$\hat{\mu}_{MEL}$	−0.01	0.31	1.00	<0.01	0.19	1.00	−0.01	0.14	1.00
(d)	$\hat{\mu}_{AIPW}$	0.08	0.38	1.47	0.10	0.27	1.95	0.10	0.20	2.12
	$\hat{\mu}_{EL}$	0.05	0.35	1.27	0.09	0.23	1.50	0.08	0.18	1.66
	$\hat{\mu}_{CAL}$	0.07	0.38	1.49	0.10	0.25	1.76	0.10	0.19	2.00
	$\hat{\mu}_{MEL}$	<0.01	0.31	1.00	0.01	0.19	1.00	0.01	0.14	1.00

formance. When the outcome regression model was misspecified, the modified empirical likelihood estimators had smaller bias and variability compared to the empirical likelihood and calibration estimators, which were consistent with the theoretical results. In this simulation study, the modified empirical likelihood estimator performed consistently better than other estimators.

The second simulation study was an adaptation of the scenario in Qin and Zhang (2007). Sample sizes for each simulated data set was 200, 500 or 1000, and 1000 Monte Carlo datasets were generated. For each observation, a standard normal random variable X was generated. Conditional on $X = x$, Y is distributed from a normal distribution with mean $2 + 3x^2$ and variance x^2. The non-missing indicator R follows a generalized linear model with a logit or a complementary log-log link:

$$\text{logit} P(R = 1|X) = 1 + 0.5X$$

or

$$\log(-\log(1 - P(R = 1|X))) = 1 + 0.5X$$

In both cases, we assumed a logistic working model for R, so that the working model is misspecified under the complementary log-log model. We considered $a(X) = X^2$ and $a(X) = X$, and the latter corresponded to a misspecified working regression model. The simulation results are shown in Table 2. Under this scenario, we had very similar conclusions as in Table 1. In particular, the modified empirical likelihood estimator performed consistently better than other

estimators, including the empirical likelihood estimator. The improvement in efficiency was greater when the working outcome regression model was misspecified.

5. Discussions

Empirical likelihood has gained increasing attention in missing data analysis since it has a good small sample performance and is also easy to compute. In this paper we studied modifications of empirical likelihood estimator that attain uniform improvement in asymptotic efficiency. Results from simulation studies were consistent with the theoretical results that the modifications improve efficiencies.

We consider the class of influence functions $\mathcal{L}$ noting that certain important existing estimators are within that class, including the empirical likelihood estimator of Qin and Zhang (2007). A different empirical likelihood estimator has been proposed by Qin, Zhang and Leung (2009) that minimizes asymptotic variance in the following class of influence functions

$$
\mathcal{L}'' = \left\{ a \left[\frac{R}{\pi_0(X)} Y - \mu \right] + b \left[\frac{R - \pi_0(X)}{\pi(X)} g(X) \right] \right. \\
\left. + c^T \frac{\partial \pi}{\partial \beta} \frac{R - \pi_0(X)}{\pi_0(X)(1 - \pi_0(X))} : a, b \in \mathbb{R}, c \in \mathbb{R}^l \right\}.
$$

where l is the dimension of β and $g(X)$ is a fixed function. The function $g(X)$ plays a similar role as $m(X)$ in $\mathcal{L}$, however, we allow $m(X)$ to be any linear combination of $\tilde{a}$ in the class $\mathcal{L}$ but $g(X)$ is fixed in advance. Another major difference is that estimators in the class $\mathcal{L}''$ may not be doubly robust when a and b are not identical. However, estimators having influence functions in class $\mathcal{L}$ enjoy double robustness properties in general. Since double robustness offer protection against misspecification of missing data model and can be useful in practice, we therefore keep our attention to the class $\mathcal{L}$.

The method of Qin, Zhang and Leung (2009) was designed for general estimating functions. A reviewer suggested us to extend the modified empirical likelihood estimation to the general case. Let $U(Y, X; \beta)$ be an unbiased estimating function. Consider the class of influence function

$$
\mathcal{Q} = \left\{ \frac{R}{\pi_0(X)} U(Y, X; \beta_0) + \frac{R - \pi_0(X)}{\pi_0(X)} m(X) : m(X) \text{ is linear in } \tilde{a}(X) \right\}
$$

where β_0 is the true value of β. Since p_i^{MEL} can be computed once $\tilde{a}(X)$ is specified, it seems reasonable to solve the weighted estimating equation

$$
\sum_{i=1}^{n} r_i p_i^{MEL} U(y_i, x_i; \beta) = 0
$$

to obtain an estimate of β. Statistical properties of the weighted estimating equation will be studied in the future.

References

CHEN, J., SITTER, R. and WU, C. (2002). Using empirical likelihood methods to obtain range restricted weights in regression estimators for surveys. *Biometrika* **89** 230–237. MR1888365

DEVILLE, J. and SÄRNDAL, (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association* **87** 376–382. MR1173804

HALL, A. and INOUE, A. (2003). The large sample behaviour of the generalized method of moments estimator in misspecified models. *Journal of Econometrics* **114** 361–394. MR1977725

KANG, J. and SCHAFER, J. (2007). Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical science* **22** 523–539. MR2420458

NEWEY, W. and MCFADDEN, D. (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics* **4** 2111–2245. MR1315971

QIN, J. and LAWLESS, J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics* **22** 300–325. MR1272085

QIN, J. and ZHANG, B. (2007). Empirical-likelihood-based inference in missing response problems and its application in observational studies. *Journal of the Royal Statistical Society: Series B* **69** 101–122. MR2301502

QIN, J., ZHANG, B. and LEUNG, D. (2009). Empirical Likelihood in Missing Data Problems. *Journal of the American Statistical Association* **104** 1492–1503. MR2750574

ROBINS, J., ROTNITZKY, A. and ZHAO, L. (1994). Estimation of Regression Coefficients When Some Regressors Are Not Always Observed. *Journal of the American Statistical Association* **89** 846–866. MR1294730

WANG, D. and CHEN, S. (2009). Empirical likelihood for estimating equations with missing values. *The Annals of Statistics* **37** 490–517. MR2488360

WANG, Q. and RAO, J. (2002). Empirical likelihood-based inference under imputation for missing response data. *Annals of statistics* 896–924. MR1922545

WHITE, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50** 1–25. MR0640163