2014

Population empirical likelihood for nonparametric inference in survey sampling

Sixia Chen
Iowa State University

Jae Kwang Kim
Iowa State University, jkim@iastate.edu

Statistica Sinica **24** (2014), 335-355
doi:http://dx.doi.org/10.5705/ss.2011.294

POPULATION EMPIRICAL LIKELIHOOD FOR NONPARAMETRIC INFERENCE IN SURVEY SAMPLING

Sixia Chen and Jae Kwang Kim

Iowa State University

Abstract: Empirical likelihood is a popular tool for incorporating auxiliary information and constructing nonparametric confidence intervals. In survey sampling, sample elements are often selected by using an unequal probability sampling method and the empirical likelihood function needs to be modified to account for the unequal probability sampling. Wu and Rao (2006) proposed a way of constructing confidence regions using the pseudo empirical likelihood of Chen and Sitter (1999).

In this paper, we propose using empirical likelihood in survey sampling based on the so-called population empirical likelihood (POEL). In the POEL approach, a single empirical likelihood is defined for the finite population. The sampling design can be incorporated into the constraint in the optimization of the POEL. For some special sampling designs, the proposed method leads to optimal estimation and does not require artificial adjustment for constructing likelihood ratio confidence intervals. Furthermore, because a single empirical likelihood is defined for the finite population, it naturally incorporates auxiliary information obtained from multiple surveys. Results from two simulation studies are presented to show the finite sample performance of the proposed method.

Key words and phrases: Calibration estimation, optimal estimation, regression estimation, Wilk's theorem.

1. Introduction

The empirical likelihood method, Owen (1988, 1990), provides a useful tool for obtaining nonparametric confidence regions for statistical functionals. Even though the empirical likelihood method is a nonparametric approach in the sense that it does not require a parametric model for the underlying distribution of the sample observations, the empirical likelihood method shares most of the desirable properties of the likelihood-based method. Using a nonparametric likelihood function, the empirical likelihood method can easily incorporate both known constraints on parameters and prior information about parameters obtained from other sources. For example, Chen and Qin (1993), Qin (2000), and Chaudhuri, Handcock, and Rendall (2008) discussed combining information using empirical likelihood. Qin and Lawless (1994) considered the situation when the parameter of interest is the solution to an estimating equation. A comprehensive overview of the empirical likelihood method is provided by Owen (2001).

When the sample is selected by an unequal probability sampling method from the finite population, the empirical likelihood needs to be modified to incorporate the sampling design. Chen and Sitter (1999) considered the pseudo empirical likelihood estimator that uses the sampling weight in the empirical log-likelihood function. Kim (2009) considered an alternative empirical likelihood function based on the biased sampling likelihood of Vardi (1985) and Qin (1993). In either case, the resulting empirical likelihood estimator naturally incorporates the available population information and achieves optimality under some limited situations. Because the empirical likelihood function is changed to incorporate the unequal probability sampling design, the resulting confidence interval based on the likelihood ratio does not have a limiting chi-square distribution and often extra computations, as discussed in Wu and Rao (2006), are required to obtain a Wilk-type confidence region. Furthermore, the sample-based empirical likelihood approach can be problematic if we want to combine information from two independent surveys, since there are different empirical likelihood functions associated with each sample.

In this paper, we propose a novel approach for the empirical likelihood in survey sampling based on the so-called population empirical likelihood (POEL). In this POEL approach, a single empirical likelihood is defined for the finite population and the sampling design can be incorporated as a constraint in the empirical likelihood. For some sampling designs, such as the Poisson sampling or the rejective Poisson sampling of Fuller (2009a), the proposed method leads to optimal estimation and the likelihood ratio follows a chi-square distribution in the limit if the sampling rate is negligible. Thus, unlike the pseudo empirical likelihood method, a Wilk-type confidence interval based on the POEL can be constructed without any artificial adjustment. Furthermore, because a single empirical likelihood is defined for the entire finite population, it naturally incorporates the setup of combining multiple surveys. The resulting empirical likelihood estimator is asymptotically equivalent to the optimal estimator obtained by the generalized method of moments (GMM), but it avoids the burden of computing the variance-covariance matrix for the GMM computation.

The rest of this paper is organized as follows. In Section 2, basic setup is introduced and the population empirical likelihood method is presented. In Section 3, some asymptotic properties of the proposed estimator are discussed under Poisson sampling. The proposed method is extended to the rejective Poisson sampling in Section 4, and extended to the problem of combining independent surveys in Section 5. Results from two limited simulation studies are presented in Section 6. Concluding remarks are made in Section 7. All technical details are given in the Supplementary Material.

2. Population Empirical Likelihood

Consider a finite population (x_i, y_i) of known size N. Suppose we are interested in estimating a parameter θ_0 that is defined by solving

$$\sum_{i=1}^{N} U(x_i, y_i; \theta) = 0, \tag{2.1}$$

for θ. Many finite-population parameters can be defined as the solutions to (2.1). Thus, if the parameter of interest is the population total $Y = \sum_{i=1}^{N} y_i$, we can take $\theta = N\mu_y$ and μ_y through (2.1) with $U(X, Y; \mu_y) = (Y - \mu_y)$. Without loss of generality, we assume that the solution θ_0 to (2.1) is unique.

Suppose now that a sample of size n is selected from the population using a probability sampling design. Let s be the index set of the sample and $\pi_i = Pr\,(i \in s)$, the known first-order inclusion probabilities of unit i, for all units in the population. Let $d_i = \pi_i^{-1}$ be the design weight of unit i in the sample. A design-consistent estimator of θ_0 can be obtained by solving the estimating equation

$$N^{-1} \sum_{i \in s} d_i U\left(x_i, y_i; \theta\right) = 0 \tag{2.2}$$

for θ. Binder (1983) discussed some theories for the estimators defined from the solution to the estimating equation (2.2) under complex sampling.

If we know aggregate population information on x, such as the population mean $\bar{X}_N$, then we can incorporate it to improve efficiency of the resulting estimator of θ_0. One way to achieve this efficiency is through calibration. Thus, instead of solving (2.2), consider solving

$$\sum_{i \in s} d_i \omega_i U\left(x_i, y_i; \theta\right) = 0, \tag{2.3}$$

where ω_i is determined to minimize $\sum_{i \in s} d_i \left(\omega_i - 1\right)^2$ subject to the calibration constraint

$$\sum_{i \in s} d_i \omega_i (1, x_i')' = (1, \bar{X}_N')'. \tag{2.4}$$

For the special case of $\theta_0 = \bar{Y}_N = N^{-1} \sum_{i=1}^{N} y_i$, Deville and Särndal (1992) discussed the choice of objective functions that lead to calibration estimators asymptotically equivalent to the generalized regression (GREG) estimator

$$\hat{\theta}_{GREG} = \bar{y}_d - \hat{B}\left(\bar{x}_d - \bar{X}_N\right), \tag{2.5}$$

where

$$\left(\bar{x}_d', \bar{y}_d\right)' = \left(\sum_{i \in s} d_i\right)^{-1} \sum_{i \in s} d_i \left(x_i', y_i\right)',$$

$$\hat{B} = \sum_{i \in s} d_i y_i \left(x_i - \bar{x}_d\right)' \left\{ \sum_{i \in s} d_i \left(x_i - \bar{x}_d\right) \left(x_i - \bar{x}_d\right)' \right\}^{-1}.$$

Chen and Sitter (1999) considered using the pseudo empirical likelihood function

$$l_p(\omega) = \sum_{i \in s} d_i \log\left(d_i \omega_i\right) \tag{2.6}$$

as an objective function for the calibration estimation with constraints (2.4). The resulting pseudo empirical likelihood calibration estimator for $\theta_0 = \bar{Y}_N$ is asymptotically equivalent to the GREG estimator in (2.5). The GREG estimator has certain optimal properties under the model where the finite population is a realization of the linear regression model

$$y_i = x_i'\beta + e_i, \tag{2.7}$$

with $E_\zeta(e_i) = 0$ and $V_\zeta(e_i) = \sigma^2$. If the linear regression model (2.7) does not hold, then the GREG estimator is no longer optimal.

The design optimal regression estimator that minimizes the design variance among the linear class $\hat{\theta} = \bar{y}_{HT} - B(\bar{q}_{HT} - \bar{q}_N)$ is

$$\hat{\theta}_{opt} = \bar{y}_{HT} - \hat{B}_{opt}\left(\bar{q}_{HT} - \bar{q}_N\right), \tag{2.8}$$

where

$$(\bar{q}_{HT}', \bar{y}_{HT})' = (N^{-1} \sum_{i \in s} d_i q_i', N^{-1} \sum_{i \in s} d_i y_i)',$$

$q_i = (1, x_i')'$, and $\hat{B}_{opt}$ is a consistent estimator of $B_{opt} = \text{Cov}\ (\bar{y}_{HT}, \bar{q}_{HT})$ $\{\text{Var}\ (\bar{q}_{HT})\}^{-1}$. The design optimal regression estimator has been discussed by Fuller and Isaki (1981), Montanari (1987), and Rao (1994).

We consider an empirical-likelihood-type estimator that leads to a solution asymptotically equivalent to the design optimal regression estimator in (2.8). Instead of assigning weights only for the sample, we propose using the population-level log-likelihood

$$l = \sum_{i=1}^{N} \log\left(\omega_i\right), \tag{2.9}$$

$\sum_{i=1}^N \omega_i = 1$, as the objective function for the calibration estimation. Because the final estimator is obtained by solving (2.3) for θ, the final weights $d_i\omega_i$ in (2.3) are used to compute the design optimal estimator from the sample observation. Unlike the pseudo empirical likelihood, the proposed likelihood (2.9), called the population empirical likelihood (POEL), is defined at the population level. To incorporate the auxiliary information into the estimation, we use

$$\sum_{i \in s} d_i \omega_i (1, x_i')' = (1, \bar{X}_N')',$$

the constraints in (2.4). For rejective Poisson sampling, in order to remove the effect of sampling design, we can incorporate additional constraints in the sampling design, as will be discussed in Section 4.

There are several advantages of the proposed method. First, it naturally incorporates additional information. For example, if $\bar{h}_N = N^{-1}\sum_{i=1}^N h(x_i)$ is known, $h(x)$ an arbitrary function of x, then we can add the constraint

$$\sum_{i\in s} d_i\omega_i h(x_i) = \bar{h}_N,$$

into the optimization using the POEL. Thus it is directly applicable in the calibration problem of survey sampling. Given the constraints, it achieves the lower bound for the asymptotic design variance under some sampling designs. For example, if $\theta_0 = \bar{Y}_N$ and $h(x) = (1, x')'$, we show that the proposed estimator is asymptotically equal to design optimal regression estimator (2.8) when the sampling rate is negligible. In addition, under some regularity conditions, the POEL enables us to obtain the likelihood ratio confidence intervals using chi-square quantiles. Then too, we can combine all sources of information from several surveys by using a single POEL to obtain the optimal estimator, as will be discussed in Section 5.

3. Main Results

Consider a Poisson sampling setup where independent Bernoulli trials are used to select the sample. Let I_i be the sample selection indicator that takes the value one if unit i is selected in the sample and takes the value zero otherwise. In the Poisson sampling, the I_i are independent Bernoulli (π_i) random variables, where π_i are known.

Under Poisson sampling, the POEL approach can be formulated as maximizing

$$l = \sum_{i=1}^N \log(\omega_i), \tag{3.1}$$

subject to

$$\sum_{i=1}^N \omega_i = 1, \quad \sum_{i=1}^N \omega_i \frac{I_i}{\pi_i} U_i(\theta) = 0. \tag{3.2}$$

Thus, without extra information, we get $\omega_i = N^{-1}$ and the POEL estimator $\hat{\theta}_{POEL}$ is the same as that obtained from the solution of (2.2). In order to incorporate the known population size information, we add the constraint

$$\sum_{i=1}^N \omega_i (\frac{I_i}{\pi_i} - 1) = 0. \tag{3.3}$$

In the constraints (3.2) and (3.3), the observed values of x_i in the units with $I_i = 0$ are not used. To incorporate the auxiliary information associated with the non-sampled part of x_i, we can impose that

$$\sum_{i=1}^{N} \omega_i \frac{I_i}{\pi_i} h_i = \bar{h}_N, \tag{3.4}$$

for some function $h_i = h(x_i)$, where $\bar{h}_N = N^{-1}\sum_{i=1}^{N} h(x_i)$. By (3.2) and (3.3), (3.4) can be written as

$$\sum_{i=1}^{N} \omega_i \frac{I_i}{\pi_i} \left(h_i - \bar{h}_N\right) = 0. \tag{3.5}$$

To solve for this optimization problem, by the Lagrange multiplier method, a two-step algorithm can be used. In the first step, the optimal weight that maximizes (3.1) subject to $\sum_{i=1}^{N} \omega_i = 1$, (3.3), and (3.5) can be expressed as

$$\hat{\omega}_i = \frac{1}{N} \frac{1}{1 + \hat{\lambda}' g_i},$$

where $g_i = ((I_i \pi_i^{-1} - 1), I_i \pi_i^{-1}(h_i - \bar{h}_N)')'$ and $\hat{\lambda}$ is the solution to

$$\frac{1}{N} \sum_{i=1}^{N} \frac{g_i}{1 + \hat{\lambda}' g_i} = 0.$$

In the second step, we can get the resulting POEL estimator $\hat{\theta}_{POEL}$ by solving

$$\sum_{i=1}^{N} \hat{\omega}_i \frac{I_i}{\pi_i} U_i(\theta) = 0. \tag{3.6}$$

Because the control function h_i in (3.4) does not depend on θ, the POEL estimator is obtained by this two-step algorithm. Such an algorithm was discussed in Chaudhuri, Handcock, and Rendall (2008). If the control function h_i depends on the unknown parameter θ, say $h_i = h(x_i; \theta)$, then the optimization is computationally more challenging. In this case, using $\hat{h}_i = h(x_i; \hat{\theta})$, $\hat{\theta}$ any $\sqrt{n}$-consistent estimator of θ, in (3.4) leads to the same two-step algorithm for optimization and the two-step solution is asymptotically equivalent to the original solution.

To discuss asymptotic properties, we first assume a sequence of finite populations and samples satisfying regularity conditions

(C1) $\theta_0 \in \Theta$ is the unique solution to $N^{-1}\sum_{i=1}^{N} U(X_i, Y_i; \theta_0) = 0$, Θ is a compact set in p-dimensional Euclidean space, and $U(X, Y; \theta)$ is uniformly continuous in Θ.

(C2) The partial derivative $\dot{U}(\theta) = \partial U(X, Y; \theta)/\partial\theta$ is a continuous function of θ in the neighborhood of θ_0 almost surely, and $\partial U(\theta_0)/\partial\theta$ is nonsingular.

(C3) With $g_i = \left(I_i\pi_i^{-1} - 1, I_i\pi_i^{-1}(h_i - \bar{h}_N)'\right)'$, as $n_B \to \infty$,

$$n_B^{1/2}\Big(N^{-1}\sum_{i=1}^{N} I_i\pi_i^{-1}U_i'(\theta_0), N^{-1}\sum_{i=1}^{N} g_i'\Big)' \xrightarrow{d} N(0, V),$$

where $n_B = E(n)$ and V is a positive definite matrix.

(C4) $||\partial U(x, y; \theta)/\partial\theta||$, $||U(x, y; \theta)||^4$, and $||h(x)||^4$ are bounded by $K(x, y)$ in Θ, and $\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N} K(x_i, y_i) = \mu_K$ where $\mu_K > 0$.

(C5) $\max_{i\in s} ||h_i|| = o_p(n_B^{1/2})$ and $\max_{i\in s} ||U_i(\theta_0)|| = o_p(n_B^{1/2})$.

(C6) $C_1 < \pi_i N n_B^{-1} < C_2, \quad i = 1, 2, \ldots, N$, for some constants $0 < C_1 < C_2$.

Condition (C1) and (C2) ensure the identifiability of parameter θ_0 and the smoothness properties of function $U(\theta)$. Condition (C3) ensures the asymptotic normality of Horvitz-Thompson type estimator under Poisson sampling. Theorem 1.3.3 of Fuller (2009b) provides sufficient conditions for (C3). Condition (C4) is the usual moment condition in survey sampling. Condition (C5) is one of the typical conditions to enable $\hat{\lambda} = O_p(n_B^{-1/2})$ and Taylor expansion, while (C6) controls the behavior of the first order inclusion probabilities.

Theorem 1. *If* (C1)−(C6) *hold, the population empirical likelihood (POEL) estimator* $\hat{\theta}_{POEL}$ *at* (3.6) *has the asymptotic expansion*

$$\hat{\theta}_{POEL} - \theta_0 = -\tau\Big\{\frac{1}{N}\sum_{i=1}^{N}\frac{I_i}{\pi_i}U_i(\theta_0) - B^*(\frac{1}{N}\sum_{i=1}^{N}\frac{I_i}{\pi_i}\eta_i - \bar{\eta}_N)\Big\} + o_p(n_B^{-1/2}), \quad (3.7)$$

where $\tau = \left[N^{-1}\sum_{i=1}^{N}\partial U_i(\theta_0)/\partial\theta\right]^{-1}$, $\eta = (1, (h - \bar{h}_N)')'$, $h = h(x)$, *and* $B^* = \Omega_1\Omega_2^{-1}$, *with*

$$\Omega_1 = \Big(\frac{1}{N^2}\sum_{i=1}^{N}(\frac{1}{\pi_i} - 1)U_i, \frac{1}{N^2}\sum_{i=1}^{N}\frac{1}{\pi_i}U_i(h_i - \bar{h}_N)'\Big), \quad (3.8)$$

$$\Omega_2 = \begin{pmatrix} N^{-2}\sum_{i=1}^{N}(\pi_i^{-1} - 1) & N^{-2}\sum_{i=1}^{N}(\pi_i^{-1} - 1)(h_i - \bar{h}_N)' \\ N^{-2}\sum_{i=1}^{N}(\pi_i^{-1} - 1)(h_i - \bar{h}_N) & N^{-2}\sum_{i=1}^{N}\pi_i^{-1}(h_i - \bar{h}_N)^{\otimes 2} \end{pmatrix}, \quad (3.9)$$

where $X^{\otimes 2} = XX'$. *We have*

$$V_h^{-1/2}\left(\hat{\theta}_{POEL} - \theta_0\right) \xrightarrow{d} N(0, I), \quad (3.10)$$

and $V_h = \tau\Omega_h\tau'$ *with*

$$\Omega_h = N^{-2} V\Big\{\sum_{i=1}^{N} \frac{I_i}{\pi_i} U_i - B^*(\sum_{i=1}^{N} \frac{I_i}{\pi_i}\eta_i - \sum_{i=1}^{N} \eta_i)\Big\}.$$

Remark 1. For $\theta_0 = \bar{Y}_N$, $h = x$, and $U = y - \theta$, (3.7) is

$$\begin{aligned}\hat{\theta}_{POEL} = \bar{Y}_N &+ \frac{1}{N}\sum_{i=1}^{N} \frac{I_i}{\pi_i}(y_i - \bar{Y}_N) - B_1^*(\frac{\hat{N}}{N} - 1)\\ &-B_2^*\{\frac{1}{N}\sum_{i=1}^{N} \frac{I_i}{\pi_i}(x_i - \bar{X}_N)\} + o_p(n_B^{-1/2}),\end{aligned} \tag{3.11}$$

where $\hat{N} = \sum_{i=1}^{N} I_i\pi_i^{-1}$, $(B_1^*, B_2^*) = \Omega_1\Omega_2^{-1}$ with Ω_1 and Ω_2 at (3.8) and (3.9). If $n_B/N \to 0$, under Poisson sampling,

$$\Omega_1 = \mathrm{Cov}\left(N^{-1}\sum_{i=1}^{N} I_i\pi_i^{-1}U_i, N^{-1}\sum_{i=1}^{N} I_i\pi_i^{-1}q_i\right) + o_p(n_B^{-1})$$

and $\Omega_2 = \mathrm{Var}\,(N^{-1}\sum_{i=1}^{N} I_i\pi_i^{-1}q_i) + o_p(n_B^{-1})$ with $q_i = (1, (x_i - \bar{X}_N)')'$. Here $\hat{\theta}_{POEL}$ is obtained by minimizing the first order asymptotic variance of the estimators in the class of (3.11). Thus, it is asymptotically equivalent to the optimal estimator (2.8).

In Theorem 1, the sampling design is not necessarily Poisson sampling, though the optimality result in Remark 1 is established under Poisson sampling. By Theorem 1, the consistent estimator of V_h can be written as $\hat{V}_h = \hat{\tau}\hat{\Omega}_h\hat{\tau}'$, where $\hat{\tau} = \left\{N^{-1}\sum_{i=1}^{N} I_i\pi_i^{-1}\partial U_i(\hat{\theta})/\partial\theta\right\}^{-1}$, $\hat{\Omega}_h = N^{-2}\sum_{i=1}^{N} I_i(1-\pi_i)\pi_i^{-2}\hat{r}_i^{\otimes 2}$, $\hat{r}_i = U_i(\hat{\theta}) - \hat{B}^*\eta_i$, $\hat{B}^* = \hat{\Omega}_1\hat{\Omega}_2^{-1}$,

$$\hat{\Omega}_1 = \big(\frac{1}{N^2}\sum_{i=1}^{N} I_i(1-\pi_i)\pi_i^{-2}U_i(\hat{\theta}), \frac{1}{N^2}\sum_{i=1}^{N} I_i\pi_i^{-2}U_i(\hat{\theta})(h_i - \bar{h}_N)'\big),$$

$$\hat{\Omega}_2 = \begin{pmatrix} N^{-2}\sum_{i=1}^{N} I_i(1-\pi_i)\pi_i^{-2} & N^{-2}\sum_{i=1}^{N} I_i(1-\pi_i)\pi_i^{-2}(h_i - \bar{h}_N)' \\ N^{-2}\sum_{i=1}^{N} I_i(1-\pi_i)\pi_i^{-2}(h_i - \bar{h}_N) & N^{-2}\sum_{i=1}^{N} I_i\pi_i^{-2}(h_i - \bar{h}_N)^{\otimes 2} \end{pmatrix},$$

with $\hat{\theta} = \hat{\theta}_{POEL}$.

By Theorem 1, we can construct a Wald-type confidence interval for θ_0 using the asymptotic normality.

Theorem 2. *Under the assumptions of Theorem 1, if* $R_n(\theta_0) = 2\{l(\hat{\theta}_{POEL}) - l(\theta_0)\}$ *where* $l(\theta) = \sum_{i=1}^{N} \log(\omega_i)$ *with* ω_i *satisfying* (3.2), (3.3), *and* (3.5), *then, as* $n_B \to \infty$ *and* $n_B/N \to 0$, $R_n(\theta_0) \xrightarrow{d} \chi_p^2$, *where* p *is the dimension of* θ.

According to Theorem 2, under some regularity conditions, a Wilk-type confidence interval for θ_0 can be constructed with a chi-square distribution as the limiting distribution when the sampling rate n_B/N is negligible.

The variance of the POEL estimator depends on the choice of the control function h_i in constraint (3.4). The optimal choice of h_i requires some superpopulation model for the conditional distribution of y_i on x_i. Because the mode of inference is purely design-based in our paper, we do not pursue this topic.

Remark 2. The sample empirical likelihood (SEL) estimator $\hat{\theta}_{SEL}$ can be obtained by maximizing $l_e = \sum_{i\in s}\log(\omega_i)$ subject to

$$\sum_{i\in s}\omega_i = 1, \quad \sum_{i\in s}\omega_i\pi_i^{-1}U_i(\theta) = 0, \tag{3.12}$$

$$\sum_{i\in s}\omega_i\pi_i^{-1}(h_i - \bar{h}_N) = 0. \tag{3.13}$$

The resulting SEL estimator is algebraically equivalent to the nonparametric likelihood estimator proposed by Kim (2009). Furthermore, under certain conditions, it can be shown that

$$R_n(\theta_0) = 2\left\{l_e(\hat{\theta}_{SEL}) - l_e(\theta_0)\right\}\xrightarrow{d}\chi_1^2.$$

For $\theta = E(Y)$, if $n/N \to 0$, the SEL estimator $\hat{\theta}_{SEL}$ with $h = x$ is asymptotically equivalent to the optimal estimator

$$\hat{\theta}_{opt1} = \bar{y}_d - \hat{B}(\bar{x}_d - \bar{X}_N), \tag{3.14}$$

where

$$(\bar{x}_d', \bar{y}_d) = \frac{\sum_{i\in s}\pi_i^{-1}x_i', \sum_{i\in s}\pi_i^{-1}y_i}{\sum_{i\in s}\pi_i^{-1}},$$

$\hat{B} = \hat{C}(\bar{y}_d, \bar{x}_d)\{\hat{V}(\bar{x}_d)\}^{-1}$, and $\hat{C}(\bar{y}_d, \bar{x}_d)$ and $\hat{V}(\bar{x}_d)$ are design consistent estimator of $\mathrm{Cov}\,(\bar{y}_d, \bar{x}_d)$ and $Var(\bar{x}_d)$, respectively. Comparing (3.14) with (2.8), the POEL estimator is more efficient than the SEL estimator.

4. Extension to Rejective Poisson Sampling

We now extend the results in Section 3 to other sampling designs. In particular, we consider rejective Poisson sampling, which covers simple random sampling and stratified random sampling as special cases. Rejective Poisson sampling has been studied by Hájek (1964), Hájek (1981), and Fuller (2009a). Hájek (1964) considered the linear design constraint

$$\sum_{i=1}^{N}\frac{\delta_i}{p_i}z_i = \sum_{i=1}^{N}z_i, \tag{4.1}$$

with $z_i = p_i$ and $\sum_{i=1}^N p_i = n$, where p_i and δ_i are the inclusion probabilities and sampling indicators for the initial sampling design, respectively. Fuller (2009a) considered a rejective sampling with constraints

$$\hat{Q}_{p,n} = (\bar{z}_p - \bar{Z}_N)' V_{\bar{z}\bar{z}}^{-1} (\bar{z}_p - \bar{Z}_N) < \gamma^2, \tag{4.2}$$

for some $\gamma^2 > 0$, where $\bar{z}_p = N^{-1} \sum_{i=1}^N \delta_i p_i^{-1} z_i$ and $V_{\bar{z}\bar{z}} = V_{poi}(\bar{z}_p)$, V_{poi} denoting the variance calculated under Poisson sampling design. Since (4.1) is a special case of (4.2), we consider only (4.2). We consider the following rejective Poisson sampling procedure.

Step 1 For $i = 1, \ldots, N$, generate $\delta_i \sim Bernoulli(p_i)$ independently.

Step 2 Check if (4.2) holds. If it does not hold, then go to [Step 1]. If the constraint is satisfied, then set $(I_1, \ldots, I_N) = (\delta_1, \ldots, \delta_N)$. The final sample consists of elements with $I_i = 1$.

Even if δ_i are generated independently, the realized sampling indicators $I_1, \ldots, I_N$ are no longer independent. The initial selection probabilities $p_i (i = 1, 2, \ldots, N)$ for Poisson sampling are not equal to the target inclusion probabilities $\pi_i (i = 1, 2, \ldots, N)$. The POEL estimator can be obtained by maximizing (3.1) subject to

$$\sum_{i=1}^N \omega_i = 1, \qquad \sum_{i=1}^N \omega_i (\frac{I_i}{p_i} - 1) = 0, \quad \sum_{i=1}^N \omega_i (\frac{I_i}{p_i} - 1) z_i = 0, \tag{4.3}$$

$$\sum_{i=1}^N \omega_i \frac{I_i}{p_i} (h_i - \bar{h}_N) = 0, \quad \sum_{i=1}^N \omega_i \frac{I_i}{p_i} U_i(\theta) = 0. \tag{4.4}$$

In (4.3), the constraint $\sum_{i=1}^N \omega_i (I_i p_i^{-1} - 1) z_i = 0$ is added to account for the design constraint in (4.2). Suppose (C1)-(C3) and (C5) hold with π_i and g_i replaced by p_i and g_i^*, respectively, where $g_i^* = ((I_i p_i^{-1} - 1) z_i^*, I_i p_i^{-1}(h_i' - \bar{h}_N'))'$, $z_i^* = (1, z_i')'$. Let $G_N(\gamma^2) = Pr(\hat{Q}_{p,n} \le \gamma^2), G_{N(i)}(\gamma^2) = Pr(\hat{Q}_{p,n} \le \gamma^2 | i \in s), G_{N(ij)}(\gamma^2) = Pr(\hat{Q}_{p,n} \le \gamma^2 | i, j \in s)$.

(C7) $|n_B N^{-1} p_i^{-1}|, z_i$ are bounded.

(C8) $||\partial M(\theta)/\partial\theta|| and ||M(\theta)||^4$ are bounded by $K(x, y)$ in Θ, and $\lim_{N\to\infty} N^{-1} \sum_{i=1}^N K(x_i, y_i) = \mu_K$ for some $\mu_K > 0$, where $M_i(\theta) = (U_i'(\theta), z_i^{*'}, h_i' - \bar{h}_N')'$.

(C9) $G_{N(i)}(\gamma^2) = G_N(\gamma^2) + g_{1N}(\gamma^2)\gamma^2 \eta_i + o_p(n_B^{-1})$, where $\eta_i = n_B N^{-2}(1 - p_i) p_i^{-1} z_i^2$ and $g_{1N}(\gamma^2)$ is a bounded sequence.

(C10) $G_{N(ij)}(\gamma^2) = G_N(\gamma^2) + g_{1N}(\gamma^2)\gamma^2(\eta_i + \eta_j) + o_p(n_B^{-1})$.

(C11) $V_{\bar{z}\bar{z}}^{-1/2}(\bar{z}_p - \bar{Z}_N) \xrightarrow{d} N(0, I)$.

Condition (C7) is used to control the behavior of the first order inclusion probabilities and the boundness of z_i. Condition (C8) is the usual moment condition in survey sampling. Conditions (C9) and (C10) are similar to Assumption 8 in Fuller (2009a), while (C11) holds for Poisson sampling under the moment conditions specified in Theorem 1.3.3 of Fuller (2009b). For motivation of (C9) and (C10), without loss of generality assume $\bar{z}_N = 0$ and $n_B N^{-2} \sum_{i=1}^{N} (1 - p_i) p_i^{-1} z_i^2 = 1$. After some algebra,

$$E(\bar{z}_p - \bar{Z}_N | i \in s) = \frac{1}{N} \frac{1 - p_i}{p_i} z_i, \tag{4.5}$$

$$E(\bar{z}_p - \bar{Z}_N | i, j \in s) = \frac{1}{N} \frac{1 - p_i}{p_i} z_i + \frac{1}{N} \frac{1 - p_j}{p_j} z_j, \tag{4.6}$$

$$Var(\bar{z}_p - \bar{Z}_N | i \in s) = n_B^{-1} - \frac{1}{N^2} \frac{1 - p_i}{p_i} z_i^2, \tag{4.7}$$

$$Var(\bar{z}_p - \bar{Z}_N | i, j \in s) = n_B^{-1} - \frac{1}{N^2} \frac{1 - p_i}{p_i} z_i^2 - \frac{1}{N^2} \frac{1 - p_j}{p_j} z_j^2. \tag{4.8}$$

According to (C11), G_N is the CDF of the Chi-square distribution and $G_{N(i)}$ and $G_{N(ij)}$ are the CDF of noncentralized Chi-square distributions. According to (4.5)-(4.8), we have

$$E(\hat{Q}_{p,n} | i \in s) = 1 - \eta_i + o_p(n_B^{-1}), \quad E(\hat{Q}_{p,n} | i, j \in s) = 1 - \eta_i - \eta_j + o_p(n_B^{-1}),$$

where $\eta_i = n_B N^{-2} (1 - p_i) p_i^{-1} z_i^2$. So, we can write

$$\begin{aligned} G_{N(i)}(\gamma^2) &= Pr(\hat{Q}_{p,n} \le \gamma^2 | i \in s) = Pr\left\{ (1 - \eta_i)^{-1} \hat{Q}_{p,n} \le (1 - \eta_i)^{-1} \gamma^2 | i \in s \right\} \\ &= G_N\left\{ (1 - \eta_i)^{-1} \gamma^2 \right\} + o_p(n_B^{-1}) \\ &= G_N((1 + \eta_i)\gamma^2) + o_p(n_B^{-1}) \\ &= G_N(\gamma^2) + g_{1N}(\gamma^2)\gamma^2 \eta_i + o_p(n_B^{-1}), \end{aligned}$$

where g_{1N} is the density of the Chi-square distribution. Similarly,

$$G_{N(ij)}(\gamma^2) = G_N(\gamma^2) + g_{1N}(\gamma^2)\gamma^2(\eta_i + \eta_j) + o_p(n_B^{-1}).$$

Theorem 3. *Consider a rejective Poisson sampling with the design constraint in* (4.2). *Let $\hat{\theta}_{POEL}$ be the population empirical likelihood estimator obtained by maximizing* (3.1) *subject to constraints* (4.3) *and* (4.4). *If* (C1)−(C3), (C5) *and* (C6), *and* (C7)−(C11) *hold,*

$$\hat{\theta}_{POEL} - \theta_0 = -\tau \Big\{ \frac{1}{N} \sum_{i=1}^{N} \frac{I_i}{\pi_i} U_i(\theta_0) - B\big(\frac{1}{N} \sum_{i=1}^{N} \frac{I_i}{\pi_i} \eta_i - \bar{\eta}_N\big) \Big\} + o_p(n_B^{-1/2}), \tag{4.9}$$

where $\tau = \left\{N^{-1}\sum_{i=1}^{N}\partial U_i(\theta_0)/\partial\theta\right\}^{-1}$, $\eta_i = (z_i^{*\prime}, (h_i - \bar{h}_N)')'$, $h_i = h(x_i)$, $z_i^* = (1, z_i')'$, $B = \Omega_1\Omega_2^{-1}$, *where*

$$\Omega_1 = \Big(\frac{1}{N^2}\sum_{i=1}^{N}(\frac{1}{\pi_i} - 1)U_i z_i^{*\prime}, \frac{1}{N^2}\sum_{i=1}^{N}\frac{1}{\pi_i}U_i(h_i - \bar{h}_N)'\Big), \tag{4.10}$$

$$\Omega_2 = \begin{pmatrix} N^{-2}\sum_{i=1}^{N}(\pi_i^{-1} - 1)z_i^{*\otimes 2} & N^{-2}\sum_{i=1}^{N}(\pi_i^{-1} - 1)z_i^*(h_i - \bar{h}_N)' \\ N^{-2}\sum_{i=1}^{N}(\pi_i^{-1} - 1)(h_i - \bar{h}_N)z_i^{*\prime} & N^{-2}\sum_{i=1}^{N}\pi_i^{-1}(h_i - \bar{h}_N)^{\otimes 2} \end{pmatrix}. \tag{4.11}$$

We have

$$V_h^{-1/2}\left(\hat{\theta}_{POEL} - \theta_0\right) \xrightarrow{d} N(0, I), \tag{4.12}$$

where $V_h = \tau\Omega_h\tau'$ *with*

$$\begin{aligned}\Omega_h &= N^{-2}V\Big\{\sum_{i=1}^{N}\frac{I_i}{\pi_i}U_i - B(\sum_{i=1}^{N}\frac{I_i}{\pi_i}\eta_i - \sum_{i=1}^{N}\eta_i)\Big\} \\ &= V_{poi}(\hat{\bar{e}}_p),\end{aligned}$$

V_{poi} *denotes the variance under Poisson sampling design, and* $\hat{\bar{e}}_p = N^{-1}\sum_{i=1}^{N} I_i p_i^{-1} e_i$ *with* $e_i = U_i - B\eta_i$.

Remark 3. For $\hat{\theta}_0 = \bar{Y}_N$ and $h = x$, (4.9) simplifies to

$$\begin{aligned}\hat{\theta}_{POEL} = \bar{Y}_N + \frac{1}{N}\sum_{i=1}^{N}\frac{I_i}{\pi_i}(y_i - \bar{Y}_N) - B_1(\frac{1}{N}\sum_{i=1}^{N}\frac{I_i}{\pi_i} - 1)z_i^* \\ -B_2\{\frac{1}{N}\sum_{i=1}^{N}\frac{I_i}{\pi_i}(x_i - \bar{X}_N)\} + o_p(n_B^{-1/2}),\end{aligned}$$

where $(B_1, B_2) = \Omega_1\Omega_2^{-1}$ with Ω_1 and Ω_2 defined at (4.10) and (4.11). If $\gamma = o(1)$ in (4.2), then

$$\bar{z}_{HT} - \bar{Z}_N = o_p(n_B^{-1/2}), \tag{4.13}$$

with $\bar{z}_{HT} = N^{-1}\sum_{i=1}^{N} I_i\pi_i^{-1}z_i$. When $n_B/N \to 0$, by (4.13), we have

$$\begin{aligned}\hat{\theta}_{POEL} = \bar{Y}_N + \frac{1}{N}\sum_{i=1}^{N}\frac{I_i}{\pi_i}(y_i - \bar{Y}_N) - B_1^*(\frac{1}{N}\sum_{i=1}^{N}\frac{I_i}{\pi_i} - 1) \\ -B_2^*\{\frac{1}{N}\sum_{i=1}^{N}\frac{I_i}{\pi_i}(x_i - \bar{X}_N)\} + o_p(n_B^{-1/2}),\end{aligned}$$

where $z_i^* = (1, z_i')'$, $(B_1^*, B_2^*) = \Omega_1^*\Omega_2^{*-1}$,

$$\Omega_1^* = N^{-2}\sum_{i=1}^N (1-\pi_i)\pi_i^{-1}(y_i - \bar{Y}_N)q_i' - \Big\{N^{-2}\sum_{i=1}^N (1-\pi_i)\pi_i^{-1}(y_i-\bar{Y}_N)z_i'\Big\}$$
$$\times\Big\{\sum_{i=1}^N (1-\pi_i)\pi_i^{-1}z_i z_i'\Big\}^{-1}\Big\{\sum_{i=1}^N (1-\pi_i)\pi_i^{-1}z_i q_i'\Big\},$$
$$\Omega_2^* = N^{-2}\sum_{i=1}^N (1-\pi_i)\pi_i^{-1}q_i q_i'$$
$$-\Big\{N^{-2}\sum_{i=1}^N (1-\pi_i)\pi_i^{-1}q_i z_i'\Big\}\Big\{\sum_{i=1}^N (1-\pi_i)\pi_i^{-1}z_i z_i'\Big\}^{-1}\{\sum_{i=1}^N (1-\pi_i)z_i q_i'\},$$

with $q_i = (1, (x_i - \bar{X}_N)')'$. Under some regularity conditions, it can be shown that $\Omega_1^* = \mathrm{Cov}\,(N^{-1}\sum_{i=1}^N I_i\pi_i^{-1}U_i, N^{-1}\sum_{i=1}^N I_i\pi_i^{-1}q_i) + o_p(n_B^{-1})$ and $\Omega_2^* = \mathrm{Var}\,(N^{-1}\sum_{i=1}^N I_i\pi_i^{-1}q_i) + o_p(n_B^{-1})$. Thus, by using a similar argument as in Remark 1, we have $\hat{\theta}_{POEL} = \hat{\theta}_{opt} + o_p(n_B^{-1/2})$, $\hat{\theta}_{opt}$ defined at (2.8).

A consistent variance estimator of $\hat{\theta}_{POEL}$ can be constructed with $\hat{V}_h = \hat{\tau}\hat{\Omega}_h\hat{\tau}'$, $\hat{\tau} = \big\{N^{-1}\sum_{i\in s} p_i^{-1}\partial U_i(\hat{\theta})/\partial\theta\big\}^{-1}$, and $\hat{\Omega}_h = N^{-2}\sum_{i=1}^N I_i(1-p_i)p_i^{-2}\hat{r}_i^{\otimes 2}$, where $\hat{r}_i = U_i(\hat{\theta}) - \hat{B}^*\eta_i$, $\hat{B}^* = \hat{\Omega}_1\hat{\Omega}_2^{-1}$ and

$$\hat{\Omega}_1 = \Big(\frac{1}{N^2}\sum_{i=1}^N I_i(1-p_i)p_i^{-2}U_i(\hat{\theta})z_i^{*'}, \frac{1}{N^2}\sum_{i=1}^N I_i p_i^{-2}U_i(\hat{\theta})(h_i - \bar{h}_N)'\Big),$$
$$\hat{\Omega}_2 = \begin{pmatrix} N^{-2}\sum_{i=1}^N I_i(1-p_i)p_i^{-2}z_i^{*\otimes 2} & N^{-2}\sum_{i=1}^N I_i(1-p_i)p_i^{-2}z_i^*(h_i-\bar{h}_N)' \\ N^{-2}\sum_{i=1}^N I_i(1-p_i)p_i^{-2}(h_i-\bar{h}_N)z_i^{*'} & N^{-2}\sum_{i=1}^N I_i p_i^{-2}(h_i - \bar{h}_N)^{\otimes 2}\end{pmatrix}.$$

Theorem 4. *Suppose the sample is obtained as a rejective Poisson sampling design and that the regularity conditions in Theorem* 3 *hold. If* $R_n(\theta_0) = 2\{l(\hat{\theta}_{POEL}) - l(\theta_0)\}$, *where* $l(\theta) = \sum_{i=1}^N \log(\omega_i)$ *with* ω_i *satisfying* (4.3) *and* (4.4), *then, as* $n_B \to \infty$ *and* $n_B/N \to 0$, $R_n(\theta_0)\xrightarrow{d}\chi_p^2$, *where* p *is the dimension of* θ.

5. Combining Information from Two Independent Surveys

Consider two independent surveys, survey 1 and survey 2, from the same finite population, and suppose the auxiliary variable x_i is observed in common. In addition, we observe (z_{1i}, z_{2i}) throughout the population, where z_{1i} is observed in the survey 1 sample and z_{2i} is observed in the survey 2 sample, and suppose that an intercept is included in z_{1i} and z_{2i}. This type of sampling design is often called a non-nested two-phase sampling design (Hidiroglou (2001)). Zieschang

(1990), Renssen and Nieuwenbroek (1997), and Merkouris (2004) considered using GREG-type estimators to combine information from different surveys. Wu (2004) considered the pseudo empirical likelihood method to solve such problems and showed that the pseudo empirical likelihood estimator is asymptotically equivalent to the GREG estimator.

We propose using the population empirical likelihood method to combine information from non-nested two-phase sampling. The proposed population-level empirical likelihood method is different from the sample-level empirical likelihood method of Owen (1988) in that we use all the available information and the proposed estimator is optimal. In addition, under some regularity conditions, we can construct likelihood ratio type confidence intervals with a chi-square limiting distribution. The proposed method can be easily extended to combining more than two surveys.

For simplicity, assume that the sampling designs in two surveys are independent Poisson sampling designs. We can easily extend our results to other sampling designs. Let I_{1i} and I_{2i} be the sample selection indicators for survey 1 and survey 2, respectively, and let π_{1i} and π_{2i} be the corresponding first order inclusion probabilities.

We are interested in estimating the general parameter at (2.1). The proposed POEL procedure for combining two surveys can be formulated as maximizing

$$l = \sum_{i=1}^{N} \log(\omega_i),$$

subject to

$$\sum_{i=1}^{N} \omega_i = 1, \quad \sum_{i=1}^{N} \omega_i (I_{1i}\pi_{1i}^{-1} - 1) z_{1i} = 0, \quad \sum_{i=1}^{N} \omega_i (I_{2i}\pi_{2i}^{-1} - 1) z_{2i} = 0,$$

$$\sum_{i=1}^{N} \omega_i (I_{1i}\pi_{1i}^{-1} - I_{2i}\pi_{2i}^{-1}) h_i = 0, \quad \sum_{i=1}^{N} \omega_i I_{2i}\pi_{2i}^{-1} U_i(\theta) = 0.$$

Under the regularity conditions of Theorem 1 for each survey, if $n/N \to 0$, it can be shown that our proposed estimator $\hat{\theta}_{POEL}$ is asymptotically equivalent to the optimal estimator that minimizes

$$Q(\bar{h}_N, \theta) = \begin{pmatrix} \bar{z}_{HT,1} - \bar{Z}_1 \\ \bar{h}_{HT,1} - \bar{h}_N \\ \bar{h}_{HT,2} - \bar{h}_N \\ \bar{z}_{HT,2} - \bar{Z}_2 \\ \bar{U}_{HT,2}(\theta) \end{pmatrix}' V_Q^{-1} \begin{pmatrix} \bar{z}_{HT,1} - \bar{Z}_1 \\ \bar{h}_{HT,1} - \bar{h}_N \\ \bar{h}_{HT,2} - \bar{h}_N \\ \bar{z}_{HT,2} - \bar{Z}_2 \\ \bar{U}_{HT,2}(\theta) \end{pmatrix}, \tag{5.1}$$

with respect to $\bar{h}_N$ and θ, where $(\bar{z}_{HT,1}, \bar{h}_{HT,1}) = N^{-1}\sum_{i=1}^{N} I_{1i}\pi_{1i}^{-1}(z_{1i}, h_i)$, $(\bar{z}_{HT,2}, \bar{h}_{HT,2}) = N^{-1}\sum_{i=1}^{N} I_{2i}\pi_{2i}^{-1}(z_{2i}, h_i)$, $(\bar{Z}_1, \bar{Z}_2) = N^{-1}\sum_{i=1}^{N}(z_{1i}, z_{2i})$, $\bar{h}_N =$

$N^{-1}\sum_{i=1}^N h_i$, $\bar{U}_{HT,2}(\theta) = N^{-1}\sum_{i=1}^N I_{i2}\pi_{2i}^{-1}U_i(\theta)$, and V_Q is the estimated variance-covariance matrix of $(\bar{z}_{HT,1}, \bar{h}_{HT,1}, \bar{h}_{HT,2}, \bar{z}_{HT,2}, \bar{U}_{HT,2}(\theta))$. The optimal estimator obtained by minimizing (5.1) is called the generalized method of moment (GMM) estimator (Hansen (1982)). The GMM estimator is a popular tool for combining information from several sources in the econometrics literature (Imbens and Lancaster (1994)); Hirano et al. (1998)). Imbens (2002) showed the asymptotic equivalence between the empirical likelihood estimator and the GMM estimator under the single sample setup. To compute the GMM estimator from (5.1), we need to estimate the variance-covariance matrix. The empirical likelihood approach avoids the computation for the variance-covariance matrix.

For the special case $\theta_0 = \bar{Y}_N$ and $h_i = x_i$, the optimal estimator of θ_0 minimizing (5.1) can be written as

$$\hat{\theta}_{opt} = \bar{y}_{HT,2} + \hat{B}_{1opt}(\bar{Z}_1 - \bar{z}_{HT,1}) + \hat{B}_{2opt}(\bar{Z}_2 - \bar{z}_{HT,2}) + \hat{B}_{3opt}(\bar{x}_{HT,1} - \bar{x}_{HT,2}),$$

where $\bar{y}_{HT,2} = N^{-1}\sum_{i=1}^N I_{2i}\pi_{2i}^{-1}y_i$,

$$\bar{x}_{HT,t} = N^{-1}\sum_{i=1}^N I_{ti}\pi_{ti}^{-1}x_i, \quad \bar{z}_{HT,t} = N^{-1}\sum_{i=1}^N I_{ti}\pi_{ti}^{-1}z_{ti}, \quad t = 1, 2,$$
$$\hat{B}_{opt} = (\hat{B}_{1opt}, \hat{B}_{1opt}, \hat{B}_{1opt}) = \hat{C}(\bar{y}_{HT,2}, \bar{S}_{HT})\{\hat{V}(\bar{S}_{HT})\}^{-1},$$

with $\bar{S}_{HT} = (\bar{z}'_{HT,1} - \bar{Z}'_1, \bar{z}'_{HT,2} - \bar{Z}'_2, \bar{x}'_{HT,2} - \bar{x}'_{HT,1})'$, $\hat{C}(\bar{y}_{HT,2}, \bar{S}_{HT})$ and $\hat{V}(\bar{S}_{HT})$ consistent estimators of $\mathrm{Cov}(\bar{y}_{HT,2}, \bar{S}_{HT})$ and $Var(\bar{S}_{HT})$, respectively. Under some regularity conditions for both surveys, we can get $2\{l(\hat{\theta}_{POEL}) - l(\theta_0)\} \xrightarrow{d} \chi_1^2$, useful for constructing a Wilk-type confidence interval.

6. Simulation Study

6.1. Simulation one

We performed two limited simulation studies. The first can be described as a $2 \times 3 \times 4$ factorial design with three factors. The first factor is the model for generating the finite population; the second is the sampling design; the third is the estimation method. Two finite populations of (x_i, y_i, z_i), population A and population B, with size $N = 10,000$ were generated. In population A, the population elements were generated by $z_i \sim \chi^2(2) + 1$, $x_i = a_i + z_i$ and $y_i = 1 + 1.2(x_i - 3) + (x_i/4)e_i$, where $a_i \sim N(0, 1)$, independent of z_i, and $e_i \sim \chi^2(1) - 1$, independent of (a_i, z_i). In population B, (x_i, z_i) were the same as in population A, and $y_i = 0.2(x_i - 1)^2 + (x_i/4)e_i$. From each population, $n = 200$ sample elements were selected repeatedly for $B = 2,000$ times. For the sampling design, three sampling designs were considered: simple random sampling (SRS) without replacement, Poisson sampling, and rejective Poisson sampling. For the

Poisson sampling, we used $\pi_i = nz_i/(\sum_{i=1}^N z_i)$. In the rejective Poisson sampling, the fixed-size constraint $\sum_{i=1}^N I_i = n$ was used with the initial sample selection probability $p_i = nz_i/(\sum_{i=1}^N z_i)$. The parameter of interest is the population mean of y. From each sample, six point estimators were computed.

1. Hájek (HJ) estimator: $\hat{\theta}_{HJ} = \sum_{i\in s} \pi_i^{-1} y_i / \sum_{i\in s} \pi_i^{-1}$.
2. Horvitz-Thompson (HT) estimator: $\hat{\theta}_{HT} = N^{-1} \sum_{i\in s} \pi_i^{-1} y_i$.
3. Proposed population-level empirical likelihood (POEL1) method without using x information, obtained by maximizing $l = \sum_{i=1}^N \log(\omega_i)$ subject to (3.2), (3.3) and $\sum_{i=1}^N \omega_i (I_i - p_i) = 0$ (for SRSWOR and rejective Poisson sampling) with $U = y - \theta$.
4. Pseudo-empirical likelihood (PEL) method with constraint (2.4).
5. Proposed sample-level empirical likelihood (SEL) method in Remark 2 by using constraints (3.12), (3.13), and design constraint (for SRSWOR and rejective Poisson sampling) $\sum_{i\in s} \omega_i p_i^{-1}(p_i - \bar{p}_N) = 0$ with $U = y - \theta$ and $h = x$.
6. Proposed population-level empirical likelihood (POEL2) method using constraints (3.2), (3.3), (3.4), and $\sum_{i=1}^N \omega_i (I_i - p_i) = 0$ (for SRSWOR and rejective Poisson sampling) with $U = y - \theta$ and $h = x$.

The first three estimators are computed without using x information while the other estimators incorporate the population mean of x. In rejective Poisson sampling, the constraint $\sum_{i=1}^N \omega_i (I_i p_i^{-1} - 1) p_i = 0$ is added to account for the design information $\sum_{i=1}^N I_i = n$. Based on $B = 2,000$ Monte Carlo samples, we have computed the biases, variances, and mean squared errors of the six estimators. Table 1 presents the results. The HJ and HT estimators are identical under SRS, but HT estimator is more efficient than the HJ estimator under other designs. The PEL estimator performs well in Case A, where the regression model holds for the finite population, because it is asymptotically equivalent to the GREG estimator. The POEL1 estimator has the same efficiency as the HJ and HT estimators under SRS, but it performs better under other designs because it effectively uses the population size (N) information. The three empirical likelihood methods (PEL, SEL, POEL2) using x information show similar performances in both populations under SRS, but SEL and POEL2 are more efficient than the PEL estimator for other designs because the SEL and POEL2 methods incorporate the design information more efficiently than the PEL method.

In addition to point estimators, we also computed interval estimators for the POEL2 method with a 95% nominal coverage. The interval estimators were computed by the likelihood ratio method based on the results in Theorem 2 and Theorem 4. Table 2 presents the simulation results of the interval estimators. In Table 2, Wald-type confidence intervals were constructed as $(\hat{\theta} - 2\sqrt{\hat{V}}, \hat{\theta} + 2\sqrt{\hat{V}})$,

Table 1. Monte Carlo biases, variances, and mean squared errors of the point estimators.

Population	Design	Method	Bias	Var	MSE
A	SRSWOR	HJ	-0.006	0.046	0.046
		HT	-0.006	0.046	0.046
		POEL1	-0.006	0.046	0.046
		PEL	-0.003	0.010	0.010
		SEL	-0.001	0.009	0.009
		POEL2	-0.001	0.009	0.009
	Poisson	HJ	0.011	0.043	0.043
		HT	0.004	0.035	0.035
		POEL1	0.004	0.035	0.035
		PEL	0.001	0.008	0.008
		SEL	0.003	0.007	0.007
		POEL2	0.003	0.007	0.007
	Rejective Poisson	HJ	0.000	0.039	0.039
		HT	-0.004	0.028	0.028
		POEL1	-0.002	0.016	0.0165
		PEL	-0.005	0.008	0.008
		SEL	-0.002	0.006	0.006
		POEL2	-0.002	0.006	0.006
B	SRSWOR	HJ	-0.005	0.070	0.070
		HT	-0.005	0.070	0.070
		POEL1	-0.005	0.070	0.070
		PEL	-0.005	0.024	0.024
		SEL	-0.003	0.024	0.024
		POEL2	-0.003	0.024	0.024
	Poisson	HJ	0.007	0.038	0.038
		HT	0.000	0.034	0.034
		POEL1	0.000	0.030	0.030
		PEL	-0.001	0.022	0.022
		SEL	-0.001	0.016	0.016
		POEL2	-0.002	0.016	0.016
	Rejective Poisson	HJ	0.003	0.037	0.037
		HT	-0.002	0.019	0.019
		POEL1	-0.003	0.014	0.014
		PEL	-0.001	0.022	0.022
		SEL	-0.004	0.013	0.013
		POEL2	-0.004	0.013	0.013

HJ: Hájek estimator, HT: Horvitz-Thompson estimator, PEL: Pseudo Empirical Likelihood estimator, SEL: Proposed sample EL estimator, POEL1: Proposed population EL estimator (without using x information), POEL2: Proposed population EL estimator incorporating x information, SRSWOR: Simple Random Sampling Without Replacement.

Table 2. Coverage rate and average length comparison for Wald's and Wilk's type 95% confidence intervals of proposed POEL2 method.

Population	Sampling design	Method	Coverage rate	Average length
A	SRSWOR	Wald	0.923	0.362
		Wilk	0.934	0.379
	Poisson	Wald	0.931	0.313
		Wilk	0.942	0.327
	Rejective Poisson	Wald	0.932	0.309
		Wilk	0.944	0.322
B	SRSWOR	Wald	0.923	0.580
		Wilk	0.938	0.598
	Poisson	Wald	0.935	0.486
		Wilk	0.944	0.503
	Rejective Poisson	Wald	0.936	0.450
		Wilk	0.949	0.471

where $\hat{V}$ was computed by the plug-in method described after Theorem 1 and Theorem 3. The Wilk-type confidence intervals were computed by the method in Theorem 2 and Theorem 4. The actual coverage rates of the Wilk-type confidence intervals are very close to the nominal coverage rates in the simulation study. In general, the Wilk-type confidence intervals show better coverage properties than the Wald-type confidence intervals in terms of coverage rates. We found similar results for the SEL method.

6.2. Simulation two

In the second simulation study, we considered combining information from the two independent surveys discussed in Section 5. In this simulation, an artificial finite population of size $N = 10,000$ was generated from

$$y_i = 1 + 0.8(z_i - 3) + 1.5x_i + (\frac{z_i}{5})e_i,$$

where the z_i were generated from $\chi^2(2) + 1$, $e_i \sim \chi^2(1) - 1$, and $x_i \sim N(2, 1)$. From the finite population, we repeatedly generated two independent samples, A_1 and A_2, with sample sizes $n_1 = 500$ and $n_2 = 200$, respectively, and $B = 2,000$ times. The sampling design for survey 1 was simple random sampling without replacement with sample size $n_1 = 500$. From the survey 1 sample, we only observe x_i. The sampling design for survey 2 was rejective Poisson sampling with fixed sample size. For the rejective Poisson sampling, we used $\pi_{i2} = n_2 z_i / \sum_{i=1}^{N} z_i$ for the initial selection probability. From the survey 2 sample, we observe x_i and y_i. The parameter of interest is the population mean of y.

From each sample pair generated as above, we computed four point estimates

Table 3. The Monte Carlo biases, variances, and the mean squared errors (MSE) of the point estimators in Simulation Two.

Method	Bias	Var	MSE
Pseudo EL	0.009	0.019	0.019
Naive Optimal	0.008	0.017	0.017
Augmented Optimal	-0.002	0.006	0.006
Proposed POEL	0.002	0.006	0.006

1. Pseudo empirical likelihood estimator (Wu (2004)), denoted $\hat{\theta}_{PEL}$, and $\hat{\theta}_{PEL} = \sum_{j\in s_2} \hat{q}_j y_j$, where $\hat{q}_j$ is obtained by maximizing $l = \sum_{i\in s_1} d_{1i}\log(p_i) + \sum_{j\in s_2} d_{2j}\log(q_j)$, subject to $\sum_{i\in s_1} p_i = \sum_{j\in s_2} q_j = 1$ and $\sum_{i\in s_1} p_i x_i = \sum_{j\in s_2} q_j x_j$.
2. The naive optimal estimator $\hat{\theta}_{opt1} = \bar{y}_{d,2} + (\bar{x}_1 - \bar{x}_{d,2})\hat{B}_{opt}$, with $\bar{x}_1 = n_1^{-1}\sum_{i\in s_1} x_i$, $(\bar{x}_{d,2}, \bar{y}_{d,2}) = (\sum_{i\in s_2}\pi_{i2}^{-1})^{-1}\sum_{i\in s_2}\pi_{i2}^{-1}(x_i, y_i)$, and $\hat{B}_{opt} = \{\hat{V}(\bar{x}_1) + \hat{V}(\bar{x}_{d,2})\}^{-1}\hat{\text{Cov}}\,(\bar{y}_{d,2}, \bar{x}_{d,2})$.
3. The augmented optimal estimator $\hat{\theta}_{opt2} = \bar{y}_{d,2} + (\bar{x}_1 - \bar{x}_{d,2})\,\hat{B}_{opt1} + (\bar{\pi}_{2N} - \bar{\pi}_{d,2})\,\hat{B}_{opt2}$, where $\hat{B}_{opt} = (\hat{B}'_{opt1}, \hat{B}'_{opt2})' = \hat{V}^{-1}(\bar{S}_d)\hat{\text{Cov}}\,(\bar{y}_{d,2}, \bar{S}_d)$, $\bar{S}_d = [(\bar{x}_{d,2} - \bar{x}_1), (\bar{\pi}_{d,2} - \bar{\pi}_{2N})]'$, $\bar{\pi}_{d,2} = \sum_{i=1}^N I_{2i}\pi_{2i}^{-1}\pi_{2i}/\sum_{i=1}^N I_{2i}\pi_{2i}^{-1}$, $\bar{\pi}_{2N} = N^{-1}\sum_{i=1}^N \pi_{2i}$.
4. The proposed POEL estimator $\hat{\theta}_{POEL}$ using constraints $\sum_{i=1}^N \omega_i = 1$, $\sum_{i=1}^N \omega_i I_{1i}\pi_{1i}^{-1} = \sum_{i=1}^N \omega_i I_{2i}\pi_{2i}^{-1} = 1, \sum_{i=1}^N \omega_i I_{1i}\pi_{1i}^{-1} x_i = \sum_{i=1}^N \omega_i I_{2i}\pi_{2i}^{-1} x_i$, and two design constraints $\sum_{i\in s_1}\omega_i = n_1/N$ and $\sum_{i=1}^N \omega_i I_{2i} = \sum_{i=1}^N \omega_i \pi_{2i}$.

The augmented optimal estimator is included to show the effect of incorporating the inclusion probability into the estimation. Table 3 presents the biases, variances, and the mean squared errors of the four point estimates. The proposed POEL estimator is more efficient than the naive optimal estimator because it incorporates additional information associated with a fixed sample size for survey 2. The performance of the augmented optimal estimator is close to the proposed POEL estimator.

7. Concluding Remarks

The objective function (2.9) can be viewed as a population-level nonparametric likelihood when the finite population is treated as a random sample from a superpopulation model. In the purely design-based approach, the superpopulation model is not assumed and the objective function in (2.9) is regarded as the negation of a distance function

$$\sum_{i=1}^N \left(\frac{1}{N}\right)\log\left(\frac{1/N}{\omega_i}\right),$$

where the distance is the Kullback-Leibler divergence from $(N^{-1},\ldots,N^{-1})$ to $(\omega_1,\ldots,\omega_N)$. The sampling design is incorporated into the constraints, rather than into the objective function for optimization, when solving the population empirical likelihood estimator. Auxiliary information for the population can also be incorporated into the constraint of the population empirical likelihood method.

The optimality of the proposed estimator holds when the sampling fraction, n/N, is negligible. If the sampling rate is not negligible then, instead of (3.4), we can use $\sum_{i=1}^{N} \omega_i(I_i/\pi_i - 1)h_i = 0$ in the constraint, as suggested by Qin, Zhang, and Leung (2009) in the context of missing data problems. In this case, the calibration condition holds only asymptotically, but not exactly. Population size N is needed to implement the population empirical likelihood method. If N is unknown, the sample empirical likelihood method discussed in Remark 2, or the new approach proposed by Berger and De La Riva Torres (2012), can be used. Further extension of the proposed method, including extension to other complex sampling designs and variable selection for calibration, is a topic of future research.

Acknowledgement

We thank two anonymous referees and an associate editor for very helpful comments. The research was partially supported by Cooperative Agreement between the USDA Natural Resources Conservation Service and the Center for Survey Statistics and Methodology at Iowa State University.

References

Binder, D. A. (1983). On the variances of asymptotically normal estimators from complex surveys. *Internat. Statist. Rev.* **51**, 279-292.

Berger, Y. G. and De La Riva Torres, O. (2012). A unified theory of empirical likelihood ratio confidence intervals for survey data with unequal probabilities and non negligible sampling fractions. *Southampton Statistical Sciences Research Institute (S3RI Methodology Working Papers).*

Chaudhuri, S., Handcock, M. S., and Rendall, M. S. (2008). Generalized linear models incorporating population level information: an empirical-likelihood-based approach. *J. Roy. Statist. Soc. Ser. B* **70**, 311-328.

Chen, J. H. and Sitter, R. R. (1999). A pseudo empirical likelihood approach to the effective use of auxiliary information in complex surveys. *Statist. Sinica* **9**, 385-406.

Chen, J. H. and Qin, J. (1993). Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. *Biometrika* **80**, 107-116.

Deville, J. C. and Särndal, C. E. (1992). Calibration estimators in survey sampling. *J. Amer. Statist. Assoc.* **87**, 376-382.

Fuller, W. A. (2009a). Some design properties of a rejective sampling procedure. *Biometrika* **96**, 933-944.

Fuller, W. A. (2009b). *Sampling Statistics.* Wiley, Hoboken, New Jersey.

Fuller, W. A. and Isaki, C. (1981). Survey design under superpopulation models. In *Current Topics in Survey Sampling.* (Edited by D. Krewski, R. Platek, and J. N. K. Rao), 199-226. Academic Press, New York.

Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *Ann. Math. Statist.* **35**, 1491-1523.

Hájek, J. (1981). *Sampling from a Finite Population.* Marcel Dekker, New York.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* **50**, 1029-1054.

Hirano, K., Imbens, G., Ridder, G. and Rubin, D. (1998). Combining panel data sets with attrition and refreshment samples. *Econometrica* **69**, 1645-1659.

Hidiroglou, M. A. (2001). Double sampling. *Survey Methodology* **27**, 143-154.

Imbens, G. W. (2002). Generalized method of moments and empirical likelihood. *J. Business & Economic Statistics* **20**, 493-506.

Imbens, G. W. and Lancaster, T. (1994). Combining micro and macro data in microeconometric models. *Rev. Economic Studies* **61**, 655-680.

Kim, J. K. (2009). Calibration estimation using empirical likelihood in survey sampling. *Statist. Sinica* **19**, 145-157.

Montanari, G. E. (1987). Post-sampling efficient Q-R prediction in large-sample surveys. *Internat. Statist. Rev.* **55**, 191-202.

Merkouris, T. (2004). Combining independent regression estimators from multiple surveys. *J. Amer. Statist. Assoc.* **99**, 1131-1139.

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

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

Owen, A. B. (2001). *Empirical Likelihood.* Chapman and Hall/CRC, New York.

Qin, J. (1993). Empirical likelihood in biased sample problems. *Ann. Statist.* **21**, 1182-1196.

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

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

Qin, J., Zhang, B. and Leung, D. (2009). Empirical likelihood in missing data problems. *J. Amer. Statist. Assoc.* **104**, 1492-1503.

Rao, J. N. K. (1994). Estimating totals and distribution functions using auxiliary information at the estimation stage. *J. Official Statist.* **10**, 153-165.

Renssen, R. H. and Nieuwenbroek, N. J. (1997). Aligning estimates for common variables in two or more sample surveys. *J. Amer. Statist. Assoc.* **92**, 368-374.

Vardi, Y. (1985). Empirical distributions in selection bias models. *Ann. Statist.* **13**, 178-203.

Wu, C. (2004). Combining information from multiple surveys through the empirical likelihood method. *Canad. J. Statist.* **32**, 15-26.

Wu, C. and Rao, J. N. K. (2006). Pseudo-empirical likelihood ratio confidence intervals for complex surveys. *Canad. J. Statist.* **34**, 359-375.

Zieschang, K. D. (1990). Sample weighting methods and estimation of totals in the consumer expenditure survey. *J. Amer. Statist. Assoc.* **85**, 986-1001.

Department of Statistics, Iowa State University, Ames, Iowa, 50011, U.S.A.

E-mail: chensixiaster@gmail.com

Department of Statistics, Iowa State University, Ames, Iowa, 50011, U.S.A.

E-mail: statistician.jkim@gmail.com

(Received November 2011; accepted December 2012)