

Empirical likelihood inference for two-sample problems

CHANGBAO WU* AND YING YAN

There exists a rich body of literature on empirical likelihood methods for two-sample problems. In this paper we focus on the simple and yet very important case of making inference on the difference of two population means using the empirical likelihood approach. Our contributions to this dynamic research topic include: (i) a weighted empirical likelihood method which not only performs well but also has a major advantage in computational simplicity; (ii) a pseudo empirical likelihood method for comparing two population means when the two samples are selected by complex surveys; (iii) two-sample empirical likelihood method with missing responses; (iv) bootstrap calibration procedures for the proposed weighted and pseudo empirical likelihood methods. Results from a limited simulation study showed that our proposed methods perform very well. The methods are also applied to a real data example on family expenditures.

AMS 2000 SUBJECT CLASSIFICATIONS: Primary 62G10; secondary 62D05.

KEYWORDS AND PHRASES: Behrens-Fisher problem, Bootstrap calibration, Case-control studies, Confidence intervals, Complex surveys, Hypothesis test, Nonparametric likelihood.

1. INTRODUCTION

Two-sample problems are commonly encountered in many areas of statistics. In classical designed experiments involving two levels of a single factor, the focus is on comparing two treatments under the controlled experimental settings. In case-control studies, two independent samples are retrospectively taken, one from the case (or disease) group and one from the control group, and the main interest is to study relationships between a disease and environmental or genetic characteristics (Qin, 1998). In observational studies, comparisons are often made between two groups defined by gender, age, ethnic backgrounds, educational levels, etc. While the settings can be extremely simple, some of the problems can be very interesting and fascinating. For example, Zhou, Gao and Hui (1997) studied the effects of two races on medical costs of patients. Their interest is whether

the average medical costs for African American patients is the same as that for white patients. In the August 22, 2011 issue of *Time* magazine, Andrea Sachs reported differences in earnings between the typical good-looking worker and the below-average-looking worker over a lifetime, and his reported results are indeed very surprising.

Suppose that Y_{11}, \dots, Y_{1n_1} and Y_{21}, \dots, Y_{2n_2} are two independent and identically distributed samples from Y_1 and Y_2 , respectively, with $E(Y_1) = \mu_1$, $E(Y_2) = \mu_2$, $Var(Y_1) = \sigma_1^2$ and $Var(Y_2) = \sigma_2^2$. The well-known Behrens-Fisher problem is to test $H_0: \mu_1 = \mu_2$ against $H_1: \mu_1 \neq \mu_2$ when both Y_1 and Y_2 are normally distributed with possibly unequal variances (Behrens, 1929; Fisher, 1935, 1939) or unknown ratio σ_1^2/σ_2^2 (Ghosh and Kim, 2001). The classic two-sample T-test under the normality assumption requires that $\sigma_1^2 = \sigma_2^2$. When both sample sizes n_1 and n_2 are large and under the null hypothesis H_0 , the test statistic $T = (\bar{Y}_2 - \bar{Y}_1)/\{S_1^2/n_1 + S_2^2/n_2\}^{1/2}$ follows approximately a standard normal distribution, where $\bar{Y}_i = n_i^{-1} \sum_{j=1}^{n_i} Y_{ij}$ and $S_i^2 = (n_i - 1)^{-1} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2$, $i = 1, 2$. However, when the sample sizes are small, normal approximation to T becomes very poor, and a known better solution is to use a t -distribution with the degree of freedom calculated by the Welch-Satterthwaite equation (Welch, 1938, 1947; Satterthwaite, 1946). When the distribution of Y_{ij} is not normal but has a known form, specific adjustment to T can be made to obtain more powerful tests. For instance, Zhou, Gao and Hui (1997) studied cases where both Y_1 and Y_2 follow log-normal distributions, and they derived new tests which showed substantial improvement over the T-test for log-normally distributed sample data.

The empirical likelihood (EL) method was first proposed by Owen (1988) as a nonparametric likelihood-based alternative approach to inference on the mean of a single population. The approach has attracted immediate attention from researchers since Owen's original work, and applications of EL have been found in many areas of statistics. One of the most significant contributions to the EL methodology was the work by Qin and Lawless (1994). They showed that side information in the form of general estimating equations can be effectively incorporated into inference through constrained maximization of the empirical likelihood function. Owen's 2001 monograph provided an excellent overview of the early developments on empirical likelihood. In recent years, there have been several new areas where the empirical likelihood method has proved to be very useful. One of

*Corresponding author.

these areas is on missing data problems; see, for instance, Wang and Rao (2001, 2002a, 2002b), Wang and Veraverbeke (2002), Wang, Linton and Hardle (2004), Qin and Zhang (2007), Wang and Dai (2008), Zhou, Wan and Wang (2008), Qin, Zhang and Leung (2009) and Wang and Chen (2009), among others. Another area with substantial new development is the use of empirical likelihood for the analysis of complex survey data; see, for example, Chen and Sitter (1999), Wu (2004, 2005), Wu and Rao (2006, 2010), Kim (2009) and Rao and Wu (2010a, 2010b), among others.

There has also been revived research on two-sample problems using the empirical likelihood method. Qin (1994) used a semi-empirical likelihood approach to inference on the difference of two population means. Jing (1995) showed that the two-sample empirical likelihood for the difference of two population means is Bartlett correctable. Qin and Zhang (1997), Qin (1998) and Zhang (2000) examined the EL methods for the two-sample problems in the context of case-control studies. Claekens, Jing, Peng and Zhou (2003) derived empirical likelihood confidence regions for the comparison distribution of two populations. They also considered ROC curves which are used to compare measurements of a diagnostic test from two populations. Zhou and Liang (2005) studied empirical likelihood-based semiparametric inference for the treatment effect in the two-sample problem with censoring. Cao and Van Keilegom (2006) developed an EL-based test on whether two populations follow the same distribution.

In this paper we focus on the simple and yet very important case of making inference on the difference of two population means using the empirical likelihood approach. Our contributions to this dynamic research topic are as follows: In Section 2, we propose a weighted empirical likelihood method which has a major advantage in computational simplicity. In Section 3, we propose a pseudo empirical likelihood method for comparing means of two finite populations when the two samples are selected by complex surveys. Two-sample empirical likelihood with missing responses is discussed in Section 4. Bootstrap calibration procedures for the weighted and pseudo empirical likelihood methods are presented in Section 5. Results from a limited simulation study are reported in Section 6. A real data example is presented in Section 7. Some additional remarks are given in Section 8.

2. WEIGHTED TWO-SAMPLE EMPIRICAL LIKELIHOOD METHOD

We start with the setting for the standard two-sample empirical likelihood and provide a detailed proof of the asymptotic χ^2 distribution of the empirical likelihood ratio statistic for the difference of two population means, which hasn't been given in detail in any of the existing papers on empirical likelihood. More importantly, the proof shows the computational difficulty involved in calculating the EL ratio statistic, which motivates the weighted EL approach we propose in section 2.2.

2.1 Standard two-sample empirical likelihood

Let Y_{11}, \dots, Y_{1n_1} and Y_{21}, \dots, Y_{2n_2} be two independent and identically distributed samples from Y_1 and Y_2 , respectively, with $E(Y_1) = \mu_1$, $E(Y_2) = \mu_2$, $Var(Y_1) = \sigma_1^2$ and $Var(Y_2) = \sigma_2^2$. Let $\theta = \mu_1 - \mu_2$ be the parameter of interest. The joint empirical log-likelihood function based on the two samples is given by

$$(1) \quad l(\mathbf{p}_1, \mathbf{p}_2) = \sum_{j=1}^{n_1} \log(p_{1j}) + \sum_{j=1}^{n_2} \log(p_{2j}),$$

where $\mathbf{p}_1 = (p_{11}, \dots, p_{1n_1})'$ and $\mathbf{p}_2 = (p_{21}, \dots, p_{2n_2})'$ are the two sets of probability measure imposed respectively over the two samples. The empirical log-likelihood ratio statistic on the parameter of interest, θ , is defined as

$$(2) \quad r(\theta) = \sum_{j=1}^{n_1} \log\{n_1 \hat{p}_{1j}(\theta)\} + \sum_{j=1}^{n_2} \log\{n_2 \hat{p}_{2j}(\theta)\},$$

where $\hat{p}_{1j}(\theta)$ and $\hat{p}_{2j}(\theta)$ maximize $l(\mathbf{p}_1, \mathbf{p}_2)$ subject to the following set of constraints:

$$(3) \quad \sum_{j=1}^{n_1} p_{1j} = 1, \quad \sum_{j=1}^{n_2} p_{2j} = 1,$$

$$(4) \quad \sum_{j=1}^{n_1} p_{1j} Y_{1j} - \sum_{j=1}^{n_2} p_{2j} Y_{2j} = \theta.$$

Let $n = n_1 + n_2$ be the combined sample size and $\theta = E(Y_1) - E(Y_2)$.

Theorem 2.1. *Suppose that $\sigma_1^2 < \infty$, $\sigma_2^2 < \infty$ and $n_1/n \rightarrow \pi \in (0, 1)$ as $n \rightarrow \infty$. Then $-2r(\theta)$ converges in distribution to a χ^2 random variable with one degree of freedom as $n \rightarrow \infty$.*

Proof. If $n_1/n \rightarrow \pi \in (0, 1)$ as $n \rightarrow \infty$, then there is no need to distinguish, for instance, among $O_p(n_1^{-1/2})$, $O_p(n_2^{-1/2})$ and $O_p(n^{-1/2})$. Let μ_0 be a fixed number such that $\mu_0 = \mu_2 + O(n^{-1/2})$. We replace constraint (4) by

$$(5) \quad \sum_{j=1}^{n_1} p_{1j} Y_{1j} = \mu_0 + \theta \quad \text{and} \quad \sum_{j=1}^{n_2} p_{2j} Y_{2j} = \mu_0.$$

The constraint (5) implies (4), but (4) does not imply (5). The newly introduced μ_0 is a nuisance parameter and will eventually be profiled. It serves as a bridge for computing the EL ratio statistic for the parameter of interest, θ . Let $r(\mu_0, \theta)$ be the empirical log-likelihood ratio statistic when (4) is replaced by (5). The initial $r(\theta)$ corresponds to $r(\hat{\mu}_0, \theta)$ where $\hat{\mu}_0$ is the maximum point of $r(\mu_0, \theta)$ with respect to μ_0 for the fixed θ .

It can be shown that \tilde{p}_{1i} and \tilde{p}_{2j} which maximize $l(\mathbf{p}_1, \mathbf{p}_2)$ subject to (3) and (5) are given by

$$(6) \quad \tilde{p}_{1j} = \frac{1}{n_1 \{1 + \lambda_1 (Y_{1j} - \mu_0 - \theta)\}}$$

and

$$\tilde{p}_{2j} = \frac{1}{n_2 \{1 + \lambda_2(Y_{2j} - \mu_0)\}},$$

where λ_1 and λ_2 are the solutions to

$$(7) \quad \frac{1}{n_1} \sum_{j=1}^{n_1} \frac{Y_{1j} - \mu_0 - \theta}{1 + \lambda_1(Y_{1j} - \mu_0 - \theta)} = 0$$

and

$$\frac{1}{n_2} \sum_{j=1}^{n_2} \frac{Y_{2j} - \mu_0}{1 + \lambda_2(Y_{2j} - \mu_0)} = 0.$$

The corresponding empirical likelihood ratio statistic is given by

$$(8) \quad r(\mu_0, \theta) = - \sum_{i=1}^{n_1} \log\{1 + \lambda_1(Y_{1i} - \mu_0 - \theta)\} - \sum_{j=1}^{n_2} \log\{1 + \lambda_2(Y_{2j} - \mu_0)\}.$$

To find the maximum point of $r(\mu_0, \theta)$ with respect to μ_0 , we set $\partial r(\mu_0, \theta) / \partial \mu_0 = 0$ which gives

$$(9) \quad n_1 \lambda_1 + n_2 \lambda_2 = 0.$$

Using standard argument as in Owen (2001, pages 219–222), we also have

$$(10) \quad \lambda_1 = \frac{1}{\sigma_1^2} (\bar{Y}_1 - \mu_0 - \theta) + o_p(n^{-1/2})$$

and

$$\lambda_2 = \frac{1}{\sigma_2^2} (\bar{Y}_2 - \mu_0) + o_p(n^{-1/2}),$$

and this leads to the solution to (9) as

$$(11) \quad \hat{\mu}_0 = \gamma(\bar{Y}_1 - \theta) + (1 - \gamma)\bar{Y}_2 + o_p(n^{-1/2}),$$

where

$$\gamma = \left(\frac{n_1}{\sigma_1^2} \right) / \left(\frac{n_1}{\sigma_1^2} + \frac{n_2}{\sigma_2^2} \right).$$

Noting that $\lambda_1(Y_{1i} - \hat{\mu}_0 - \theta) = o_p(1)$ and $\lambda_2(Y_{2j} - \hat{\mu}_0) = o_p(1)$ uniformly over all i and j and applying a Taylor series expansion to (8) at $\mu_0 = \hat{\mu}_0$, we have the following asymptotic expansion to the EL ratio function:

$$\begin{aligned} -2r(\theta) &= -2r(\hat{\mu}_0, \theta) \\ &= \frac{n_1}{\sigma_1^2} (\bar{Y}_1 - \hat{\mu}_0 - \theta)^2 + \frac{n_2}{\sigma_2^2} (\bar{Y}_2 - \hat{\mu}_0)^2 + o_p(1) \\ &= (\bar{Y}_1 - \bar{Y}_2 - \theta)^2 / \left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \right) + o_p(1). \end{aligned}$$

It immediately follows that $-2r(\theta)$ converges in distribution to a χ^2 random variable with one degree of freedom. \square

A $(1 - \alpha)$ -level confidence interval on θ can be constructed as

$$(12) \quad \mathcal{C}_1 = \{\theta \mid -2r(\theta) \leq \chi_1^2(\alpha)\},$$

where $\chi_1^2(\alpha)$ is the upper $(100\alpha)\%$ quantile from the χ_1^2 distribution. The lower and upper bounds of the interval \mathcal{C}_1 requires a bi-section search algorithm. This is a computationally challenging task, because for every selected grid point on θ , one needs to maximize the EL ratio function over the nuisance parameter, μ_0 , and there is no-closed form solution to the maximum point $\hat{\mu}_0$ for any given θ .

The computational difficulties under the standard two-sample EL formulation are due to the fact that the involved Lagrange multipliers, which are determined through the set of equations (7), have to be computed based on two separate samples with an added nuisance parameter μ_0 . Such difficulties can be avoided through an alternative formulation of the EL function, for which computation procedures are virtually identical to those for one-sample EL problems.

2.2 Weighted two-sample empirical likelihood

There exist different versions of weighting schemes for the empirical likelihood function. For instance, Ren (2008) discussed a weighted empirical likelihood approach for some two-sample semiparametric models with various types of censored data, where the weights depend on the censoring type of individual observations. The method we use here is to put a weight on each sample, which is related to the sample size. This idea was first used by Fu, Wang and Wu (2008) when they discussed inferences with multiple samples.

We define the weighted empirical log-likelihood function as

$$(13) \quad l_w(\mathbf{p}_1, \mathbf{p}_2) = \frac{w_1}{n_1} \sum_{j=1}^{n_1} \log(p_{1j}) + \frac{w_2}{n_2} \sum_{j=1}^{n_2} \log(p_{2j}),$$

where $w_1 = w_2 = 1/2$. This choice of w_1 and w_2 is to facilitate the reformulation of normalization constraints (3) and the parameter constraint (4) into the alternative forms (14) and (15), to be specified below. Let the weighted EL ratio statistic $r_w(\theta)$ be defined in the same way as $r(\theta)$ of Section 2.1 under the same constraints (3) and (4) but replacing $l(\mathbf{p}_1, \mathbf{p}_2)$ by the weighted version $l_w(\mathbf{p}_1, \mathbf{p}_2)$. We have the following result concerning the asymptotic distribution of $r_w(\theta)$.

Theorem 2.2. *Suppose that $\sigma_1^2 < \infty$, $\sigma_2^2 < \infty$ and $n_1/n \rightarrow \pi \in (0, 1)$ as $n \rightarrow \infty$. Then $-2r_w(\theta)/c_1$ converges in distribution to a χ^2 random variable with one degree of freedom as $n \rightarrow \infty$, where c_1 is a scaling constant and is specified in (17).*

Proof. The most crucial step we use to prove the result is a reformulation of the constraints, which also characterizes the required computational procedures for our proposed

weighted EL. First, we note that the two normalization constraints in (3) are equivalent to $\sum_{j=1}^{n_1} p_{1j} = 1$ and $\sum_{i=1}^2 w_i \times \sum_{j=1}^{n_i} p_{ij} = 1$. Second, the constraint (4) induced by the parameter θ can be re-written as $w_1 \sum_{j=1}^{n_1} p_{1j}(Y_{1j}/w_1) + w_2 \sum_{j=1}^{n_2} (-Y_{2j}/w_2) = \theta$. The two sets of constraints (3) and (4) can now be equivalently written as

$$(14) \quad \sum_{i=1}^2 w_i \sum_{j=1}^{n_i} p_{ij} = 1,$$

$$(15) \quad \sum_{i=1}^2 w_i \sum_{j=1}^{n_i} p_{ij} \mathbf{u}_{ij} = \mathbf{0},$$

where $\mathbf{u}_{ij} = \mathbf{Z}_{ij} - \boldsymbol{\eta}$, $\mathbf{Z}_{1j} = (1, Y_{1j}/w_1)'$, $\mathbf{Z}_{2j} = (0, -Y_{2j}/w_2)'$, $\boldsymbol{\eta} = (w_1, \theta)'$. It should be noted that the choice $w_1 = w_2 = 1/2$ is important for the above reformulation of constraints. Using the standard Lagrange multiplier method, it can be shown that the p_{ij} which maximize $l_w(\mathbf{p}_1, \mathbf{p}_2)$ subject to (14) and (15) are given by $p_{ij} = 1/\{n_i(1 + \boldsymbol{\lambda}'\mathbf{u}_{ij})\}$, and the Lagrange multiplier $\boldsymbol{\lambda}$ is the solution to

$$(16) \quad g_1(\boldsymbol{\lambda}) = \sum_{i=1}^2 \frac{w_i}{n_i} \sum_{j=1}^{n_i} \frac{\mathbf{u}_{ij}}{1 + \boldsymbol{\lambda}'\mathbf{u}_{ij}} = \mathbf{0}.$$

Substituting $1/(1 + \boldsymbol{\lambda}'\mathbf{u}_{ij}) = 1 - \boldsymbol{\lambda}'\mathbf{u}_{ij}/(1 + \boldsymbol{\lambda}'\mathbf{u}_{ij})$ into (16), we have

$$\left(\sum_{i=1}^2 \frac{w_i}{n_i} \sum_{j=1}^{n_i} \frac{\mathbf{u}_{ij}\mathbf{u}'_{ij}}{1 + \boldsymbol{\lambda}'\mathbf{u}_{ij}} \right) \boldsymbol{\lambda} = \sum_{i=1}^2 \frac{w_i}{n_i} \sum_{j=1}^{n_i} \mathbf{u}_{ij}.$$

Noting that

$$\mathbf{U} = \sum_{i=1}^2 \frac{w_i}{n_i} \sum_{j=1}^{n_i} \mathbf{u}_{ij} = (0, \bar{Y}_1 - \bar{Y}_2 - \theta)',$$

and using a similar argument as in Owen (2001, pages 219–222), we get

$$\boldsymbol{\lambda} = \mathbf{D}^{-1}\mathbf{U} + o_p(n^{-1/2}),$$

where

$$\mathbf{D} = \sum_{i=1}^2 \frac{w_i}{n_i} \sum_{j=1}^{n_i} \mathbf{u}_{ij}\mathbf{u}'_{ij}$$

is a 2×2 matrix. The weighted two-sample empirical log-likelihood ratio statistics for θ can now be expanded as follows:

$$\begin{aligned} -2r_w(\theta) &= 2 \sum_{i=1}^2 \frac{w_i}{n_i} \sum_{j=1}^{n_i} \log(1 + \boldsymbol{\lambda}'\mathbf{u}_{ij}) \\ &= 2 \sum_{i=1}^2 \frac{w_i}{n_i} \sum_{j=1}^{n_i} \left(\boldsymbol{\lambda}'\mathbf{u}_{ij} - \frac{1}{2} \boldsymbol{\lambda}'\mathbf{u}_{ij}\mathbf{u}'_{ij}\boldsymbol{\lambda} \right) + o_p(n^{-1}) \end{aligned}$$

$$\begin{aligned} &= \mathbf{U}'\mathbf{D}^{-1}\mathbf{U} + o_p(n^{-1}) \\ &= d^{(22)}(\bar{Y}_1 - \bar{Y}_2 - \theta)^2 + o_p(n^{-1}), \end{aligned}$$

where $d^{(22)}$ is the second diagonal element of \mathbf{D}^{-1} . If we let

$$(17) \quad c_1 = d^{(22)} \left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right),$$

then it follows immediately that $-2r_w(\theta)/c_1$ converges in distribution to a χ^2 random variable with one degree of freedom when $\theta = E(Y_1) - E(Y_2)$. \square

The weighted two-sample EL formulation is computationally friendly, since it does not involve any nuisance parameters, and the involved equation (16) for the Lagrange multiplier can be solved using the one-sample EL algorithm (Wu, 2004). The scaling constant c_1 involves the unknown θ and can be consistently estimated by \hat{c}_1 when θ is estimated by $\hat{\theta} = \bar{Y}_1 - \bar{Y}_2$. The resulting $-2r_w(\theta)/\hat{c}_1$ still has an asymptotic χ^2_1 distribution. Consequently, a $(1-\alpha)$ -level confidence interval on θ can be constructed as

$$(18) \quad \mathcal{C}_2 = \{ \theta \mid -2r_w(\theta)/\hat{c}_1 \leq \chi^2_1(\alpha) \}.$$

The scaling constant c_1 can also be bypassed through a bootstrap calibration method. See Section 5 for further detail.

3. TWO-SAMPLE PSEUDO EMPIRICAL LIKELIHOOD FOR COMPLEX SURVEY DATA

In this section, we use a pseudo empirical likelihood (PEL) approach to develop a method which is suitable for two-sample problems involving complex survey data. The PEL method was first proposed by Chen and Sitter (1999) for point estimation, and was later further developed by Wu and Rao (2006) for conducting hypothesis tests or constructing confidence intervals based on a single complex survey sample.

Let s_i be the set of sampled units selected from the i th finite population, $i = 1, 2$. Let n_1 and n_2 be the respective sample size. Let $\pi_j^{(i)} = P(j \in s_i)$ be the inclusion probabilities, $j = 1, \dots, n_i$ and $i = 1, 2$. Let $d_{ij} = 1/\pi_j^{(i)}$ be the basic design weights and

$$\tilde{d}_{ij}(s_i) = d_{ij} / \sum_{k \in s_i} d_{ik}$$

be the normalized design weights. We follow design-based framework for inference, where the value Y_{ij} for the study variable Y , attached to the j th unit in population i , is treated as fixed, and randomization is induced by the random selection of sampled units. Let μ_1 and μ_2 be the respective finite population means. Let $\theta = \mu_1 - \mu_2$ be the parameter of interest. The two-sample pseudo empirical likelihood

function is defined as

$$(19) \quad l_{pel}(\mathbf{p}_1, \mathbf{p}_2) = w_1 \sum_{j=1}^{n_1} \tilde{d}_{1j}(s_1) \log(p_{1j}) \\ + w_2 \sum_{j=1}^{n_2} \tilde{d}_{2j}(s_2) \log(p_{2j}),$$

where $w_1 = w_2 = 1/2$. When the selection probabilities $\pi_j^{(i)}$ are all equal for the given population i , the normalized design weights $\tilde{d}_{ij}(s_i)$ reduces to $1/n_i$. In this case our proposed pseudo empirical likelihood function $l_{pel}(\mathbf{p}_1, \mathbf{p}_2)$ reduces to the weighted EL function $l_w(\mathbf{p}_1, \mathbf{p}_2)$ that has been discussed in Section 2.2. The maximum PEL estimator of θ is computed as $\hat{\theta}_{pel} = \sum_{j=1}^{n_1} \hat{p}_{1j} Y_{1j} - \sum_{j=1}^{n_2} \hat{p}_{2j} Y_{2j}$, where the \hat{p}_{ij} 's maximize $l_{pel}(\mathbf{p}_1, \mathbf{p}_2)$ subject to the normalization constraints (3), and the estimator is given by $\hat{\theta}_{pel} = \hat{\mu}_1 - \hat{\mu}_2$, where $\hat{\mu}_i = \sum_{j=1}^{n_i} \tilde{d}_{ij}(s_i) Y_{ij} = \sum_{j=1}^{n_i} d_{ij} Y_{ij} / \sum_{j=1}^{n_i} d_{ij}$ is the Hájek estimator of the finite population mean, μ_i . Under complex survey designs, the estimator $\hat{\theta}_{pel}$ is approximately design-unbiased, while the regular estimator $\hat{\theta} = \bar{Y}_1 - \bar{Y}_2$ is biased.

Let $\hat{p}_{ij}(\theta)$ be the maximizer of $l_{pel}(\mathbf{p}_1, \mathbf{p}_2)$ subject to the normalization constraints (3) and the constraint (4) induced by the parameter of interest, θ . It follows from similar arguments in Section 2.2 that $\hat{p}_{ij}(\theta) = \tilde{d}_{ij}(s_i) / (1 + \boldsymbol{\lambda}' \mathbf{u}_{ij})$, where the \mathbf{u}_{ij} is defined in Section 2.2, both $\boldsymbol{\lambda}$ and \mathbf{u}_{ij} depend on θ , and the Lagrange multiplier $\boldsymbol{\lambda}$ is the solution to

$$(20) \quad g_2(\boldsymbol{\lambda}) = \sum_{i=1}^2 w_i \sum_{j=1}^{n_i} \frac{\tilde{d}_{ij}(s_i) \mathbf{u}_{ij}}{1 + \boldsymbol{\lambda}' \mathbf{u}_{ij}} = \mathbf{0}.$$

Noting that $\hat{p}_{ij}(\hat{\theta}_{pel}) = \tilde{d}_{ij}(s_i)$, the two-sample pseudo empirical log-likelihood ratio statistic for θ is given by

$$r_{pel}(\theta) = l_{pel} \left\{ \hat{\mathbf{p}}_1(\theta), \hat{\mathbf{p}}_2(\theta) \right\} - l_{pel} \left\{ \hat{\mathbf{p}}_1(\hat{\theta}_{pel}), \hat{\mathbf{p}}_2(\hat{\theta}_{pel}) \right\} \\ = - \sum_{i=1}^2 w_i \sum_{j=1}^{n_i} \tilde{d}_{ij}(s_i) \log(1 + \boldsymbol{\lambda}' \mathbf{u}_{ij}).$$

Let $\mathbf{K} = \sum_{i=1}^2 w_i \sum_{j=1}^{n_i} \tilde{d}_{ij}(s_i) \mathbf{u}_{ij} \mathbf{u}_{ij}'$, which is a 2×2 matrix involving the unknown θ , and

$$c_2 = k^{(22)} \left\{ V \left(\sum_{j=1}^{n_1} \tilde{d}_{1j}(s_1) Y_{1j} \right) + V \left(\sum_{j=1}^{n_2} \tilde{d}_{2j}(s_2) Y_{2j} \right) \right\},$$

where $k^{(22)}$ is the second diagonal element of \mathbf{K}^{-1} , and $V(\cdot)$ refers to design-based variance. The following result can now be established using arguments similar to the proof of Theorem 2.2. Details are omitted.

Theorem 3.1. *Under the asymptotic framework and the regularity conditions C1–C3 described in Wu and Rao (2006), applicable to both populations and sampling designs,*

and assume that $n_1/n \rightarrow \pi \in (0, 1)$ as $n \rightarrow \infty$, the adjusted two-sample pseudo empirical likelihood ratio statistic $-2r_{pel}(\theta)/c_2$ converges in distribution to a χ^2 random variable with one degree of freedom as $n \rightarrow \infty$, where $\theta = \mu_1 - \mu_2$.

The definition of the scaling constant c_2 involves the unknown θ and variances of the two Hájek estimators of the population means. Let \hat{c}_2 be a consistent estimator of c_2 . The $(1 - \alpha)$ -level PEL confidence interval on θ can be constructed as

$$(21) \quad \mathcal{C}_3 = \{ \theta \mid -2r_{pel}(\theta)/\hat{c}_2 \leq \chi_1^2(\alpha) \}.$$

In Section 5, we propose a two-sample bootstrap calibration procedure which can bypass the need for calculating the scaling constant under certain sampling designs.

4. TWO-SAMPLE EMPIRICAL LIKELIHOOD WITH MISSING DATA

In this section, we extend the discussion to two-sample problems where measures on certain auxiliary variables are available for all units in both samples but values of the response variable are subject to missingness. Suppose that $\{(Y_{ij}, \mathbf{x}_{ij}), j = 1, \dots, n_i\}$ is a conceptual random sample from the i th population, $i = 1$ and 2 , where the dimensionality and the components in the vector of auxiliary variables \mathbf{x}_{ij} can be different for the two populations but measures on \mathbf{x}_{ij} are observed for all j . The response variable Y_{ij} could be missing. Let $\delta_{ij} = 1$ if Y_{ij} is observed, and $\delta_{ij} = 0$ if Y_{ij} is missing. The two actual samples may be represented by $\{(Y_{ij}, \mathbf{x}_{ij}, \delta_{ij}), j = 1, \dots, n_i, i = 1, 2\}$. We assume that $P(\delta_{ij} = 1 \mid Y_{ij}, \mathbf{x}_{ij}) = P(\delta_{ij} = 1 \mid \mathbf{x}_{ij})$. That is, the responses are missing-at-random. We further assume that the following linear regression models hold:

$$(22) \quad Y_{ij} = \mathbf{x}_{ij}' \boldsymbol{\beta}_i + \epsilon_{ij}, \quad j = 1, \dots, n_i, \quad i = 1, 2,$$

where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are the regression parameters for the two populations, and the ϵ_{ij} 's are independent error terms with mean 0 and unknown variance τ_i^2 . The parameter of interest is still $\theta = \mu_1 - \mu_2$, where $\mu_i = E(Y_{ij})$ is the unconditional mean response for the i th population.

It should be noted that when complete responses are available, measures of the auxiliary variables \mathbf{x}_{ij} provide no additional information for the unconditional population means μ_1 and μ_2 . When the Y_{ij} 's are subject to missingness, information collected on \mathbf{x}_{ij} becomes valuable and can be used to impute the missing responses through the regression model (22). Let

$$\tilde{\boldsymbol{\beta}}_i = \left(\sum_{j=1}^{n_i} \delta_{ij} \mathbf{x}_{ij} \mathbf{x}_{ij}' \right)^{-1} \sum_{j=1}^{n_i} \delta_{ij} \mathbf{x}_{ij} Y_{ij}$$

be the ordinary least square estimator of β_i using available sample data. Let

$$\tilde{Y}_{ij} = \delta_{ij}Y_{ij} + (1 - \delta_{ij})\mathbf{x}'_{ij}\tilde{\beta}_i, \quad j = 1, \dots, n_i, \quad i = 1, 2.$$

Note that $\tilde{Y}_{ij} = Y_{ij}$ is the observed response if $\delta_{ij} = 1$, and $\tilde{Y}_{ij} = \mathbf{x}'_{ij}\tilde{\beta}_i$ is the regression imputed value for Y_{ij} if $\delta_{ij} = 0$.

We consider the standard two-sample empirical likelihood formulation described in Section 2.1. Let $\tilde{r}(\theta)$ be defined in the same way as $r(\theta)$ given by (2), with constraint (4) being replaced by

$$(23) \quad \sum_{i=1}^{n_1} p_{1i}\tilde{Y}_{1i} - \sum_{j=1}^{n_2} p_{2j}\tilde{Y}_{2j} = \theta.$$

We assume that $(Y_{ij}, \mathbf{x}_{ij}, \delta_{ij})$, $j = 1, \dots, n_i$ can be viewed as an independent and identically distributed sample from $(Y_i, \mathbf{X}_i, \delta_i)$, $i = 1, 2$.

Theorem 4.1. *Suppose that the two regression models specified by (22) hold, and the error variances τ_i^2 are finite. Assume that $E(\|\mathbf{X}_i\|^2) < \infty$, $i = 1, 2$, $n_1/n \rightarrow \pi \in (0, 1)$ as $n \rightarrow \infty$. Then $-2\tilde{r}(\theta)/c_3$ converges in distribution to a χ^2 random variable with one degree of freedom as $n \rightarrow \infty$, where $\theta = \mu_1 - \mu_2$ and c_3 is a scaling constant and is specified in (24).*

Proof. The imputation procedure used here is the same as the one used by Wang and Rao (2002b) for a single sample. The major technical difficulty is that the \tilde{Y}_{ij} 's are not independent due to the use of $\tilde{\beta}_i$ for all imputed values. Under the assumed conditions and following the same arguments used by Wang and Rao (2002b), we have

$$\frac{1}{\sqrt{n_i}} \sum_{j=1}^{n_i} (\tilde{Y}_{ij} - \mu_i) \xrightarrow{d} N(0, V_i)$$

and

$$\frac{1}{n_i} \sum_{j=1}^{n_i} (\tilde{Y}_{ij} - \mu_i)^2 \xrightarrow{p} U_i,$$

where

$$\begin{aligned} V_i &= S_{i1} + S'_{i2}S_{i3}^{-1}S_{i2}\tau_i^2 + \beta'_i S_{i4}\beta_i \\ &\quad - 2S'_{i5}\beta_i\mu_i + \mu_i^2 + 2S'_{i2}S_{i3}^{-1}S_{i6}\tau_i^2, \\ U_i &= S_{i1} + \beta'_i S_{i4}\beta_i - 2S'_{i5}\beta_i\mu_i + \mu_i^2, \end{aligned}$$

and $S_{i1} = E\{\delta_i(Y_i - \mathbf{X}'_i\beta_i)^2\}$, $S_{i2} = E\{(1 - \delta_i)\mathbf{X}_i\}$, $S_{i3} = E(\delta_i\mathbf{X}_i\mathbf{X}'_i)$, $S_{i4} = E(\mathbf{X}_i\mathbf{X}'_i)$, $S_{i5} = E(\mathbf{X}_i)$, $S_{i6} = E(\delta_i\mathbf{X}_i)$. If we define

$$(24) \quad c_3 = \left(\frac{V_1}{n_1} + \frac{V_2}{n_2} \right) / \left(\frac{U_1}{n_1} + \frac{U_2}{n_2} \right),$$

then it follows from similar arguments used in the proof of Theorem 2.1 of Section 2.1 that $-2\tilde{r}(\theta)/c_3$ converges in

distribution to a χ^2 random variable with one degree of freedom. The details can be found in Yan (2010) and are omitted here. \square

In practice, the scaling constant c_3 needs to be estimated in order to use the above result for testing hypotheses or constructing confidence intervals on θ . This can be done by using plug-in moment estimators for S_{ik} , μ_i , $k = 1, \dots, 6$, $i = 1, 2$, the least square estimators of β_i , and the estimator of τ_i^2 from the fitted residuals. It is also possible to derive a similar result under the weighted empirical likelihood formulation of Section 2.2. or a result in the same spirit of Theorem 3.1. when the two samples are selected by complex survey designs and the responses are subject to missingness. These scenarios will not be discussed further in the current paper.

5. BOOTSTRAP PROCEDURES

Empirical likelihood-based hypothesis tests or confidence intervals on θ rely on the asymptotic distribution of the empirical likelihood ratio statistic. For single samples and when the sample sizes are not large, bootstrap calibration methods often provide improved performance under standard settings. For the weighted or the pseudo empirical likelihood methods for two-sample problems, there is an added value to the bootstrap methods: the scaling constants c_1 or c_2 , which depends on unknown population quantities and needs to be estimated, can be bypassed. Simulation studies also show that the bootstrap calibrated empirical likelihood methods provide improved performance for samples with small or moderate sizes.

We now present a bootstrap calibration procedure for the two-sample pseudo empirical likelihood method described in Section 3. The proposed procedure also covers the weighted empirical likelihood approach of Section 2.2 as a special case with equal probability selection of units. Our proposed procedure is modified from the one-sample bootstrap method described in Wu and Rao (2010). The two-sample pseudo empirical likelihood ratio confidence interval (24) on θ , which involves the scaling constant c_2 and the upper α -quantile from a χ^2_1 distribution, can be replaced by $\{\theta \mid r_{pel}(\theta) > b_\alpha\}$, where b_α is the α -quantile of the sampling distribution of the pseudo empirical likelihood ratio statistic, $r_{pel}(\theta)$. The value of b_α in (25) can be approximated through the following bootstrap calibration procedure. Two important ingredients of the proposed bootstrap procedure are (i) the basic design weights d_{ij} need to be treated as part of the sample data; and (ii) a bootstrap version of the two-sample pseudo empirical likelihood function should be used.

1. Select a bootstrap sample s_i^* of size n_i from the original sample s_i using simple random sampling with replacement and denote the bootstrap sample data by $\{(d_{ij}^*, Y_{ij}^*), j \in s_i^*\}$, $i = 1, 2$.

2. Let the bootstrap version of the two-sample pseudo empirical likelihood function be defined as

$$l_{pel}^*(\mathbf{p}_1, \mathbf{p}_2) = w_1 \sum_{j \in s_1^*} \tilde{d}_{1j}^*(s_1^*) \log(p_{1j}) + w_2 \sum_{j \in s_2^*} \tilde{d}_{2j}^*(s_2^*) \log(p_{2j}),$$

where $\tilde{d}_{ij}^*(s_i^*) = d_{ij}^* / \sum_{j \in s_i^*} d_{ij}^*$.

3. Calculate the bootstrap version of the two-sample pseudo empirical likelihood ratio statistic as

$$r_{pel}^*(\hat{\theta}_{pel}) = w_1 \sum_{j \in s_1^*} \tilde{d}_{1j}^*(s_1^*) \log\left(\frac{p_{1j}}{\tilde{d}_{1j}^*(s_1^*)}\right) + w_2 \sum_{j \in s_2^*} \tilde{d}_{2j}^*(s_2^*) \log\left(\frac{p_{2j}}{\tilde{d}_{2j}^*(s_2^*)}\right),$$

where the p_{ij} 's maximize $l_{pel}^*(\mathbf{p}_1, \mathbf{p}_2)$ subject to

$$\sum_{j=1}^{n_1} p_{1j} = 1, \quad \sum_{j=1}^{n_2} p_{2j} = 1, \\ \sum_{j=1}^{n_1} p_{1j} Y_{1j}^* - \sum_{j=1}^{n_2} p_{2j} Y_{2j}^* = \hat{\theta}_{pel},$$

where $\hat{\theta}_{pel} = \hat{\mu}_1 - \hat{\mu}_2$ and $\hat{\mu}_i = \sum_{j=1}^{n_i} \tilde{d}_{ij}^*(s_i^*) Y_{ij}^*$.

4. Repeat Steps 1, 2 and 3 a large number of times, B , independently, to obtain the sequence $r_1^*(\hat{\theta}_{pel}), \dots, r_B^*(\hat{\theta}_{pel})$. Let b_α^* be the 100α th sample quantile from this sequence.

The bootstrap calibrated two-sample pseudo empirical likelihood ratio confidence interval on θ can now be constructed as

$$(25) \quad \mathcal{C}_3^* = \{\theta \mid r_{pel}(\theta) > b_\alpha^*\}.$$

The following theorem states that this interval has the desired level of coverage probability. The proof is a combination of arguments used in the proof of Theorem 3.1 and those used in the proof of Theorem 1 of Wu and Rao (2010). Details are omitted.

Theorem 5.1. *The bootstrap calibrated two-sample pseudo empirical likelihood ratio confidence interval \mathcal{C}_3^* on θ has asymptotically correct coverage probability at $(1 - \alpha)$ -level if the conditions of Theorem 3.1 hold and the original samples s_1 and s_2 are selected using unequal probability sampling with replacement.*

In practice, complex survey samples are usually selected by without-replacement sampling procedures. The interval \mathcal{C}_3^* can be used for without-replacement sampling designs if the sampling fraction $f_i = n_i/N_i$ is negligible. Here N_i is the size of the i th finite population. When f_i is not small, the

interval \mathcal{C}_3^* under the proposed with-replacement bootstrap procedure tends to have an over-coverage problem, as shown by the one-sample simulation results reported in Wu and Rao (2010).

6. SIMULATION STUDIES

In this section we report results from a limited simulation study on the finite sample performance of the proposed empirical likelihood ratio confidence intervals on the difference of two population means, with comparison to the conventional T-test based method. We consider three cases: (i) Both Y_1 and Y_2 are normally distributed; (ii) Both Y_1 and Y_2 follow lognormal distributions; (iii) The two samples are selected from two finite populations, with the response variable containing many zero values. More scenarios of simulations, not reported here to save space, can be found in Yan (2010).

For cases (i) and (ii), two independent and identically distributed samples of sizes n_1 and n_2 are drawn respectively from Y_1 and Y_2 , and four confidence intervals on $\theta = \mu_1 - \mu_2$ are computed: (1) the conventional normal-approximation interval based on $T = \{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)\} / \{S_1^2/n_1 + S_2^2/n_2\}^{1/2}$ (T-Test); (2) the standard two-sample empirical likelihood method described in Section 2.1 (EL); (3) the weighted empirical likelihood method introduced in Section 2.2 (WEL); and (4) the bootstrap-calibrated weighted empirical likelihood method (BWEL). The nominal value of the confidence level is fixed at 95%. Performances of these intervals are evaluated based on coverage probability (CP), lower (L) and upper (U) tail error rates and average length (AL), computed as follows:

$$L = 100 \times \frac{1}{B} \sum_{b=1}^B I(\hat{\theta}_L^{(b)} \geq \theta),$$

$$U = 100 \times \frac{1}{B} \sum_{b=1}^B I(\hat{\theta}_U^{(b)} \leq \theta),$$

$$CP = 100 \times \frac{1}{B} \sum_{b=1}^B I(\hat{\theta}_L^{(b)} < \theta < \hat{\theta}_U^{(b)})$$

and

$$AL = \frac{1}{B} \sum_{b=1}^B (\hat{\theta}_U^{(b)} - \hat{\theta}_L^{(b)}),$$

where $(\hat{\theta}_L^{(b)}, \hat{\theta}_U^{(b)})$ is a confidence interval on θ computed from the b th simulated sample, and $B = 2,000$ is the total number of simulation runs. For the bootstrap-calibrated method, the number of bootstrap samples used for computing the interval is 1,000. It should be noted that $L + CP + U = 100$ for any method.

Table 1 summarizes the results for case (i), where $Y_1 \sim N(\mu_1, \sigma_1^2)$ and $Y_2 \sim N(\mu_2, \sigma_2^2)$, $\mu_1 = \mu_2 = 1$, $\sigma_1 = 1.5$ and $\sigma_2 = 1$. We have the following major observations: (a) The

Table 1. Confidence Intervals on $\theta = \mu_1 - \mu_2$ for Two Normal Populations

(n_1, n_2)	CI	L	CP	U	AL
(30,30)	T-Test	2.10	95.05	2.85	1.31
	EL	2.45	94.40	3.15	1.28
	WEL	2.20	94.90	2.90	1.30
	BWEL	2.05	95.25	2.70	1.31
(30,60)	T-Test	4.70	91.20	4.10	1.06
	EL	3.10	94.20	2.70	1.18
	WEL	2.75	94.60	2.65	1.20
	BWEL	2.70	94.75	2.55	1.21
(60,30)	T-Test	0.80	98.20	1.00	1.21
	EL	2.55	94.90	2.55	1.04
	WEL	2.20	95.55	2.25	1.05
	BWEL	2.40	95.10	2.50	1.05
(60,60)	T-Test	2.40	94.95	2.65	0.92
	EL	2.50	94.60	2.90	0.91
	WEL	2.45	94.70	2.85	0.92
	BWEL	2.55	94.85	2.60	0.92
(30,90)	T-Test	4.65	89.85	5.50	0.95
	EL	2.95	94.45	2.60	1.14
	WEL	2.75	94.75	2.50	1.16
	BWEL	2.60	94.90	2.50	1.16
(90,30)	T-Test	1.00	97.95	1.05	1.16
	EL	2.35	94.60	3.05	0.94
	WEL	2.30	94.70	3.00	0.95
	BWEL	2.35	94.65	3.00	0.95

Table 2. Confidence Intervals on $\theta = \mu_1 - \mu_2$ for Two Lognormal Populations

(n_1, n_2)	CI	L	CP	U	AL
(30,30)	T-Test	1.05	93.10	5.85	2.65
	EL	2.55	91.85	5.60	2.68
	WEL	2.40	92.25	5.35	2.73
	BWEL	1.50	93.75	4.75	3.07
(30,60)	T-Test	2.90	89.45	7.65	2.06
	EL	2.20	93.20	4.60	2.50
	WEL	1.65	93.15	5.20	2.55
	BWEL	0.95	94.40	4.65	2.77
(60,30)	T-Test	1.00	97.35	2.55	2.52
	EL	2.95	93.45	3.60	2.13
	WEL	3.25	94.50	2.25	2.16
	BWEL	2.50	95.65	1.85	2.32
(60,60)	T-Test	1.25	93.15	5.60	1.87
	EL	2.80	92.60	4.60	1.92
	WEL	2.65	93.00	4.35	1.94
	BWEL	2.00	94.05	3.95	2.08
(30,90)	T-Test	4.90	86.10	9.00	1.82
	EL	2.55	92.20	5.25	2.42
	WEL	1.75	91.85	6.40	2.48
	BWEL	1.10	93.05	5.85	2.64
(90,30)	T-Test	0.15	97.75	2.10	2.46
	EL	2.35	93.75	3.90	1.89
	WEL	3.15	94.60	2.25	1.92
	BWEL	2.28	95.00	2.20	2.03

T-test method provides excellent results for scenarios where $n_1 = n_2$. Coverage probabilities are very close to the nominal value, and the two tail error rates are balanced; (b) When the two sample sizes n_1 and n_2 are not equal, the T-test method performs poorly, and the interval is either too wide or too narrow, depending on which sample is bigger; (c) The empirical likelihood-based method is not sensitive to the unequal sample sizes, and the weighted EL performs uniformly better than the standard two-sample EL method. The coverage probabilities are good, the two tail error rates are very balanced, and the average length is not inflated for all cases; (d) The bootstrap-calibrated weighted EL method also provides excellent results.

Results for case (ii) where $Y_1 \sim \text{Lognormal}(\nu_1, \sigma_1^2)$, $Y_2 \sim \text{Lognormal}(\nu_2, \sigma_2^2)$, $\nu_1 = 1.1$, $\nu_2 = 1.2$, $\sigma_1^2 = 0.4$ and $\sigma_2^2 = 0.2$ are summarized in Table 2. All major points observed from case (i) still hold, except that the performances of the EL-based method seem to have deteriorated a little bit for the case $n_1 = 30$ and $n_2 = 90$. The bootstrap-calibrated EL method provides acceptable results for all cases.

For case (iii), we consider two finite populations of sizes $N_1 = N_2 = 5,000$. For the first population, $M_1 = 3,000$ responses Y_{1j} are set to zero, and for the second population, $M_2 = 4,000$ responses Y_{2j} are zero. The nonzero responses are generated from $\text{Uniform}(0.8, 1.2)$ and $\text{Uniform}(1.8, 2.2)$, respectively. The two samples are taken by simple random sampling without replacement. Under such sampling designs, the pseudo-EL method of Section 3 reduces to the weighted EL method of Section 2.2. The re-

Table 3. Confidence Intervals on $\theta = \mu_1 - \mu_2$ for Two Finite Populations

(n_1, n_2)	CI	L	CP	U	AL
(30,30)	T-Test	4.40	94.25	1.35	0.68
	EL	3.10	94.70	2.20	0.66
	WEL	2.70	95.30	2.00	0.67
	BWEL	2.30	95.65	2.05	0.68
(30,60)	T-Test	2.25	97.15	0.60	0.63
	EL	2.95	94.45	2.60	0.53
	WEL	2.20	94.95	2.85	0.54
	BWEL	2.15	94.90	2.95	0.54
(60,30)	T-Test	6.70	90.00	3.30	0.54
	EL	3.40	94.25	2.35	0.61
	WEL	4.30	93.80	1.90	0.62
	BWEL	3.10	95.05	1.85	0.63
(60,60)	T-Test	3.20	95.35	1.45	0.48
	EL	2.15	95.25	2.60	0.47
	WEL	2.05	95.55	2.40	0.48
	BWEL	2.00	95.45	2.55	0.48
(30,90)	T-Test	1.00	98.75	0.25	0.62
	EL	2.15	95.40	2.45	0.48
	WEL	1.60	95.05	3.35	0.49
	BWEL	1.60	94.90	3.50	0.48
(90,30)	T-Test	7.65	86.65	5.70	0.49
	EL	2.85	94.55	2.60	0.59
	WEL	4.00	94.10	1.90	0.60
	BWEL	2.60	95.35	2.05	0.61

sults are reported in Table 3. The T-test method once again is very sensitive to the unequal sample sizes and performs

Table 4. Confidence Intervals on the Differences in Family Expenditures

Response	CI	D_{12}	D_{13}	D_{23}
Clothing (Y)	T-Test	(-0.194, 0.364)	(1.138, 1.594)	(1.073, 1.490)
	EL	(-0.197, 0.366)	(1.146, 1.605)	(1.081, 1.506)
	WEL	(-0.203, 0.362)	(1.134, 1.597)	(1.073, 1.498)
	BWEL	(-0.207, 0.365)	(1.123, 1.607)	(1.081, 1.489)
Total (Z)	T-Test	(-7.680, 0.497)	(21.614, 28.623)	(25.744, 31.676)
	EL	(-7.630, 0.566)	(21.696, 28.732)	(25.757, 31.704)
	WEL	(-7.661, 0.541)	(21.563, 28.637)	(25.686, 31.657)
	BWEL	(-7.700, 0.581)	(21.437, 28.760)	(25.682, 31.661)

poorly with unequal sample sizes. All three EL-based methods, however, provide good and robust results for all scenarios considered in the simulation.

7. A REAL DATA EXAMPLE

We now apply the proposed weighted EL method to analyzing the data from the 2000 Statistics Canada's Family Expenditure Survey for the province of Ontario. The data set consists of $N = 2,248$ observations on variables related to family compositions, income and expenditures, including Y : the annual expenditure on clothing and Z : the total annual expenditure. One of the research questions is to see how family expenditures are related to the number of children (age ≤ 15) in the family. The original survey used a stratified simple random sampling design, but unfortunately the strata weight information is not available. In what follows, we ignore the stratification and treat the data as if they were collected by simple random sampling.

We break the data set into three subsets, corresponding to families with two or more children, one child or no children. The subsample sizes for the three groups are $n_1 = 428$, $n_2 = 579$ and $n_3 = 1,241$, respectively. For a given response variable (Y or Z), let μ_i , $i = 1, 2, 3$, be the unknown population means. We are interested in the pairwise differences $D_{ii'} = \mu_i - \mu_{i'}$. For each difference, we compute 95% confidence intervals using the four methods described in Section 6: T-Test, EL, WEL and BWEL. The results are given in Table 4 (values in \$1,000's). It is a bit surprising to see that all four methods provide very similar results. This is partially due to the relatively large sample sizes for all the three groups. Confidence intervals for D_{12} contain 0, which implies that there is no significant difference in spending between families with one child and families with two or more children. Families with one or more children spend at least \$1,000 more on clothing and at least \$21,000 more in total expenditure than families with no children.

8. CONCLUDING REMARKS

We proposed a weighted EL method for two independent and identically distributed samples and a pseudo empirical likelihood method when the two samples are taken from two

finite populations using complex survey designs. One of the major advantages of our proposed methods is the computational simplicity. Confidence intervals can be constructed using algorithms developed for one-sample problems and no nuisance parameters are involved. Under the scenarios examined in the simulation study, our proposed methods provide better results than the T-test method or the standard two-sample EL method. We are currently studying the performance of the pseudo EL method for unequal probability multistage sampling designs, as well as methods for missing data problems.

The empirical likelihood method is usually effective and powerful in dealing with populations with skewed distributions. For one sample problems, Chen, Chen and Rao (2003) and Chen and Qin (2003) applied the empirical likelihood method to populations with non-negative responses and a large portion of zero values. They showed that the empirical likelihood method is extremely efficient. The same scenario also applies to two-sample problems. Hallstrom (2010) developed a modified Wilcoxon test for comparing two populations with non-negative distributions and clumps of zeros, where parametric approaches are problematic. Our proposed two-sample empirical likelihood methods can be directly applied to such scenarios and our limited simulation results show that the EL-based methods can be very promising.

ACKNOWLEDGEMENT

This research was supported by grants to C. Wu from the Natural Sciences and Engineering Research Council of Canada and Mathematics of Information Technology and Complex Systems. The authors thank the Editor, an Associate Editor and a reviewer for many helpful comments which led to improved presentation of the paper.

Received 3 October 2011

REFERENCES

- BEHRENS, B. V. (1929). Ein Beitrag zur Fehlerberechnung bei wenige Beobachtungen (A contribution to error estimation with few observations). *Landwirtschaftliche Jahrbücher*, **68**, 807–837.
- CAO, R. and VAN KEILEGOM, I. (2006). Empirical likelihood tests for two-sample problems via nonparametric density estimation. *The Canadian Journal of Statistics*, **34**, 61–77. [MR2267710](#)

- CHEN, J. and SITTER, R. R. (1999). A pseudo empirical likelihood approach to the effective use of auxiliary information in complex surveys. *Statistica Sinica*, **9**, 385–406. [MR1707846](#)
- CHEN, J., CHEN, S. Y. and RAO, J. N. K. (2003). Empirical likelihood confidence intervals for the mean of a population containing many zero values. *The Canadian Journal of Statistics*, **31**, 53–68. [MR1985504](#)
- CHEN, S. X. and QIN, J. (2003). Empirical likelihood-based confidence intervals for data with possible zero observations. *Statistics and Probability Letters*, **65**, 29–37. [MR2012622](#)
- CLAESKENS, G., JING, B. Y., PENG, L. and ZHOU, W. (2003). Empirical likelihood confidence regions for comparison distributions and ROC curves. *The Canadian Journal of Statistics*, **31**, 173–190. [MR2016226](#)
- FISHER, R. A. (1935). The fiducial argument in statistical inference. *Annals of Eugenics*, **6**, 391–398.
- FISHER, R. A. (1939). The comparison of samples with possibly unequal variances. *Annals of Eugenics*, **9**, 174–180.
- FU, Y., WANG, X. and WU, C. (2008). Weighted empirical likelihood inference for multiple samples. *Journal of Statistical Planning and Inference*, **139**, 1462–1473. [MR2485139](#)
- GHOSH, M. and KIM, Y. (2001). The Behrens-Fisher problem revisited: a Bayes-frequentist synthesis. *The Canadian Journal of Statistics*, **29**, 5–17. [MR1834483](#)
- HALLSTROM, A. P. (2010). A modified Wilcoxon test for non-negative distributions with a clump of zeros. *Statistics in Medicine*, **29**, 391–400. [MR2750556](#)
- JING, B. Y. (1995). Two-sample empirical likelihood method. *Statistics and Probability Letters*, **24**, 315–319. [MR1353889](#)
- KIM, J. K. (2009). Calibration estimation using empirical likelihood in survey sampling. *Statistica Sinica*, **19**, 145–158. [MR2487882](#)
- OWEN, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75**, 237–249. [MR0946049](#)
- OWEN, A. B. (2001). *Empirical Likelihood*, Chapman and Hall/CRC.
- QIN, J. (1994). Semi-parametric likelihood ratio confidence intervals for the difference of two sample means. *Annals of the Institute of Statistical Mathematics*, **46**, 117–126. [MR1272752](#)
- QIN, J. (1998). Inferences for case-control and semiparametric two-sample density ratio models. *Biometrika*, **85**, 619–630. [MR1665814](#)
- QIN, J. and LAWLESS, J. F. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics*, **22**, 300–325. [MR1272085](#)
- QIN, J. and ZHANG, B. (1997). A goodness-of-fit test for logistic regression models based on case-control data. *Biometrika*, **84**, 609–618. [MR1603924](#)
- 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. H. Y. (2009). Empirical likelihood in missing data problem. *Journal of the American Statistical Association*, **104**, 1492–1503. [MR2750574](#)
- RAO, J. N. K. and WU, C. (2010a). Pseudo empirical likelihood inference for multiple frame surveys. *Journal of the American Statistical Association*, **105**, 1494–1503. [MR2796566](#)
- RAO, J. N. K. and WU, C. (2010b). Bayesian pseudo empirical likelihood intervals for complex surveys. *Journal of the Royal Statistical Society, Series B*, **72**, 533–544. [MR2758527](#)
- REN, J. J. (2008). Weighted empirical likelihood in some two-sample semiparametric models with various types of censored data. *The Annals of Statistics*, **36**, 147–166. [MR2387967](#)
- SATTERTHWAITE, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, **2**, 110–114.
- WANG, D. and CHEN, S. X. (2009). Empirical likelihood for estimating equations with missing values. *The Annals of Statistics*, **37**, 490–517. [MR2488360](#)
- WANG, L. and VERAVERBEKE, N. (2002). Empirical likelihood in a semi-parametric model for missing response data. *Communications in Statistics – Theory and Methods*, **35**, 625–639. [MR2282879](#)
- WANG, Q. and DAI, P. (2008). Semiparametric model-based inference in the presence of missing responses. *Biometrika*, **95**, 721–734. [MR2443186](#)
- WANG, Q. and RAO, J. N. K. (2001). Empirical likelihood for linear regression models under imputation for missing responses. *The Canadian Journal of Statistics*, **29**, 597–608. [MR1888507](#)
- WANG, Q. and RAO, J. N. K. (2002a). Empirical likelihood-based inference under imputation with missing response. *The Annals of Statistics*, **30**, 896–924. [MR1922545](#)
- WANG, Q. and RAO, J. N. K. (2002b). Empirical likelihood-based inference in linear models with missing data. *Scandinavian Journal of Statistics*, **29**, 563–576. [MR1925575](#)
- WANG, Q., LINTON, O. and HARDLE, W. (2004). Semiparametric regression analysis with missing response at random. *Journal of the American Statistical Association*, **99**, 334–345. [MR2062820](#)
- WELCH, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika*, **29**, 350–62.
- WELCH, B. L. (1947). The generalization of “student’s” problem when several different population variances are involved. *Biometrika*, **34**, 28–35. [MR0019277](#)
- WU, C. (2004). Some algorithmic aspects of the empirical likelihood method in survey sampling. *Statistica Sinica*, **14**, 1057–1067. [MR2126339](#)
- WU, C. (2005). Algorithms and R codes for the pseudo empirical likelihood methods in survey sampling. *Survey Methodology*, **31**, 239–243.
- WU, C., and RAO, J. N. K. (2006). Pseudo-empirical likelihood ratio confidence intervals for complex surveys. *The Canadian Journal of Statistics*, **34**, 359–375. [MR2328549](#)
- WU, C., and RAO, J. N. K. (2010). Bootstrap procedures for the pseudo empirical likelihood method in sample surveys. *Statistics and Probability Letters*, **80**, 1472–1478. [MR2669748](#)
- YAN, Y. (2010). Empirical likelihood inference for two-sample problems. Unpublished master’s thesis, Department of Statistics and Actuarial Science, University of Waterloo, Canada.
- ZHANG, B. (2000). Estimating the treatment effect in the two-sample problem with auxiliary information. *Nonparametric Statistics*, **12**, 377–389. [MR1760714](#)
- ZHOU, X. H., GAO, S. and HUI, S. L. (1997). Methods for comparing the means of two independent log-normal samples. *Biometrics*, **53**, 1129–1135.
- ZHOU, Y. and LIANG, H. (2005). Empirical likelihood-based semiparametric inference for the treatment effect in the two-sample problem with censoring. *Biometrika*, **92**, 271–282. [MR2201359](#)
- ZHOU, Y., WAN, A. T. K. and WANG, X. (2008). Estimating equations inference with missing data. *Journal of the American Statistical Association*, **103**, 1187–1199. [MR2462892](#)

Changbao Wu
 Department of Statistics and Actuarial Science
 University of Waterloo
 Waterloo, ON N2L 3G1
 Canada
 E-mail address: cbwu@uwaterloo.ca

Ying Yan
 Department of Statistics and Actuarial Science
 University of Waterloo
 Waterloo, ON N2L 3G1
 Canada
 E-mail address: yytoto@gmail.com