Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT®

Statistics & Probability Letters 66 (2004) 67–79

STATISTICS & PROBABILITY LETTERS

www.elsevier.com/locate/stapro

Weighted empirical likelihood inference

Changbao Wu

Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ont., Canada N2L 3G1

Abstract

A weighted empirical likelihood approach is proposed to take account of the heteroscedastic structure of the data. The resulting weighted empirical likelihood ratio statistic is shown to have a limiting chisquare distribution. A limited simulation study shows that the associated confidence intervals for a population mean or a regression coefficient have more accurate coverage probabilities and more balanced two-sided tail errors when the sample size is small or moderate. The proposed weighted empirical likelihood method also provides more efficient point estimators for a population mean in the presence of side information. Large sample resemblances between the weighted and the unweighted empirical likelihood estimators are characterized through high-order asymptotics and small sample discrepancies of these estimators are investigated through simulation. The proposed weighted approach reduces to the usual unweighted empirical likelihood method under a homogeneous variance structure.

Keywords: Confidence interval; Finite population; Heteroscedasticity; Linear regression model; Minimum entropy distance; Point estimation

1. Introduction

The empirical likelihood method first proposed by Owen (1988) is a powerful nonparametric inference tool with applications in many areas of statistics. New development is still active trying to extend the method to handle various non-regular situations. Owen (2001) provides a comprehensive account and an updated overview of the subject.

Suppose $Y_1, Y_2, \ldots, Y_n$ are independent and identically distributed random variables from an unknown distribution $F(y)$. Let $p_i = F(Y_i) - F(Y_i-)$. The empirical log-likelihood function $l(F) = \sum_{i=1}^n \log(p_i)$ is maximized, subject to the normalization constraints $p_i \geqslant 0$ and $\sum_{i=1}^n p_i = 1$, by the empirical distribution function, i.e. $p_1 = \cdots = p_n = n^{-1}$. Let $\mu_0 = \int y \, \mathrm{d}F(y)$ be the unknown population mean. The nonparametric maximum empirical likelihood estimator for μ_0 is given by

E-mail address: cbwu@uwaterloo.ca (C. Wu).

0167-7152/$ - see front matter
doi:10.1016/j.spl.2003.10.007

$\hat{\mu} = n^{-1} \sum_{i=1}^{n} Y_i$. Let $l(\hat{\mu}) = \sum_{i=1}^{n} \log(n^{-1})$ and $l(\mu) = \sum_{i=1}^{n} \log(p_i)$ where the p_i maximize $l(F)$ subject to

$$p_i \geqslant 0, \sum_{i=1}^{n} p_i = 1 \quad \text{and} \quad \sum_{i=1}^{n} p_i Y_i = \mu.$$

It was shown by Owen (1988) that, under some mild finite moment conditions, $-2[l(\mu_0) - l(\hat{\mu})]$ converges to χ_1^2 in distribution, and consequently a $1 - \alpha$ confidence interval for μ_0 can be constructed as

$$\{\mu \mid -2[l(\mu) - l(\hat{\mu})] < \chi_1^2(\alpha)\}, \tag{1.1}$$

where $\chi_1^2(\alpha)$ is the $1 - \alpha$ quantile from a χ^2 distribution with one degree of freedom.

More generally, if the parameter of interest, θ, is defined through a one-dimensional estimating equation $E\{U(Y, \theta)\}=0$ (or $E\{U(x, Y, \theta)\}=0$ where x is a covariate and the expectation is conditional on the given x), the foregoing formulation of (1.1) for μ_0 can be adapted for θ by simply replacing the constraint $\sum_{i=1}^{n} p_i Y_i = \mu$ by $\sum_{i=1}^{n} p_i U(Y_i, \theta) = 0$.

The empirical likelihood confidence interval (1.1) is similar to the one based on Wilks's theorem under a parametric model, and is more applicable due to its non-parametric nature and weak assumptions. There are several possible directions to extend or generalize this result. For instance, one can assume $Y_1, \ldots, Y_n$ being independent, having a common mean μ_0 but with non-constant variances $V(Y_i) = v_i \sigma^2$. In this case the confidence interval (1.1) can still be justified under slightly different moment conditions. A more general case is covered by the triangular array empirical likelihood theorem of Owen (2001).

While (1.1) has approximate confidence level of $1 - \alpha$ for large samples, its finite sample performance depends on the actual underlying distribution and cannot exactly be quantified. Some empirical evidences, however, do indicate that the actual coverage probability under small samples is usually lower than the nominal value, resulting in a false claimed confidence level. See the simulation results reported in Owen (1988) and also those reported in Section 3 of this article.

In this article we propose to use a weighted empirical likelihood approach assuming a non-constant variance structure of the data. The resulting weighted empirical likelihood ratio statistic is shown to have a limiting chisquare distribution, and the associated confidence intervals for a population mean or a regression coefficient are shown to have more accurate coverage probabilities and more balanced two-sided tail errors when the sample size is small or moderate. The proposed approach also provides more efficient point estimator for a population mean (in terms of smaller variance or mean square error) when the mean values of auxiliary variables are known. The proposed approach reduces to the usual unweighted empirical likelihood method under homogeneous variances.

Formulation of a weighted empirical likelihood function and establishment of the limiting distribution of the weighted empirical likelihood ratio statistic are presented in Section 2. The weighted and the unweighted empirical likelihood ratio confidence intervals for a regression coefficient are compared in Section 3 through a simulation study. In Section 4 the proposed weighted empirical likelihood method is applied to obtain more efficient point estimator of a population mean in the presence of side information. Large sample resemblances between the weighted and the unweighted empirical likelihood estimators are characterized through high order asymptotics and small sample

discrepancies of these estimators are investigated through simulation. All proofs are deferred to the appendix. Some additional remarks are given in Section 5.

2. Weighted empirical likelihood

Consider situations where $Y_1,\ldots,Y_n$ are independent, have a common mean μ_0 and variances $V(Y_i)=v_i\sigma^2$ where the v_i are known constants. Typically the v_i are related to certain covariates under the context of regression analysis. See Sections 3 and 4 for further illustration. Our goal is to formulate a weighted likelihood function using the v_i, an idea similar to the one used in weighted regression analysis.

The task would be much easier if we work with the so-called Euclidean likelihood function $l_E(F)=-(1/2)\sum_{i=1}^n (np_i-1)^2$ discussed in Owen (2001). This likelihood function is derived based on the Euclidean distance $\sum_{i=1}^n (p_i-n^{-1})^2$ between the two sets of probability measure $F=(p_1,\ldots,p_n)$ and $\hat{F}=(n^{-1},\ldots,n^{-1})$. A natural weighting scheme using the v_i would be $l_E^*(F)=-C_n\sum_{i=1}^n v_i(p_i-n^{-1})^2$, where $C_n>0$ is a scaling constant and its role will be clarified shortly. Such a weighting scheme reflects the relevance of the data: the larger the value of v_i, the less informative the observation Y_i, and the consequence from maximizing $l_E^*(F)$ under various constraints will force p_i taking values closer to the basic probability measure n^{-1}. This argument will become more convincing in Section 4 when a connection between the estimation of the population mean and the estimation of the underlying regression coefficients is observed. Note that maximizing $l_E^*(F)$ subject to $\sum_{i=1}^n p_i=1$ gives $p_1=\cdots=p_n=n^{-1}$.

It is not obvious that one can re-weight the empirical likelihood function $l(F)=\sum_{i=1}^n \log(p_i)$ in a similar fashion as in the case of Euclidean likelihood. Note that the empirical loglikelihood ratio statistic $r(F)=l(F)-l(\hat{F})=\sum_{i=1}^n \log(np_i)$ is not a true distance measure between $F=(p_1,\ldots,p_n)$ and $\hat{F}=(n^{-1},\ldots,n^{-1})$. Our proposed strategy is to re-weight an empirical likelihood-based distance measure using the so-called minimum entropy distance given by

$$D(F,\hat{F})=-\sum_{i=1}^n\left\{\frac{1}{n}\log(np_i)-p_i+\frac{1}{n}\right\}.$$

It can be seen that $D(F,\hat{F})$ is a true distance measure and is clearly originated from the log-likelihood function. If we impose the normalization constraint $\sum_{i=1}^n p_i=1$, then $D(F,\hat{F})=-r(F)/n$. Maximizing $r(F)$ is equivalent to minimizing $D(F,\hat{F})$. This entropy distance has previously been used in information theory as well as in survey sampling for the construction of calibration estimators. See, for instance, Deville and Särndal (1992) for further discussion.

Definition 1. Let $v_1,\ldots,v_n$ be a set of known positive constants. The weighted empirical (log) likelihood function is given by $l_W(F)=C_n\sum_{i=1}^n v_i\{\log(p_i)-np_i\}$, where $C_n=n/\sum_{i=1}^n v_i$ is a scaling constant.

One can alternatively view $l_W(F)$ as obtained by modifying $l(F)=\sum_{i=1}^n \log(p_i)$ through penalizing (the term $-np_i$) and re-weighting (the factor v_i). The very crucial nature of this formulation, however, lies in the fact that some of the basic properties of the empirical likelihood method will be preserved, as shown by Theorem 1 below and the results presented in Section 4. When the v_i are

all equal, the weighted $l_W(F)$ reduces to the unweighted $l(F)$ if one ignores a trivial constant term. To maximize $l_W(F)$ subject to $p_i > 0$ and $\sum_{i=1}^n p_i = 1$, a Lagrange multiplier argument shows that $p_1 = \cdots = p_n = n^{-1}$.

We consider a parameter θ defined through $E\{U(Y,\theta)\} = 0$ (or $E\{U(x,Y,\theta)\} = 0$). Let $l_W(\hat{\theta}) = C_n \sum_{i=1}^n v_i\{\log(n^{-1}) - 1\}$ and $l_W(\theta) = C_n \sum_{i=1}^n v_i\{\log(p_i) - n p_i\}$, where the p_i maximize $l_W(F)$ subject to $p_i > 0$, $\sum_{i=1}^n p_i = 1$ and $\sum_{i=1}^n p_i U(Y_i,\theta) = 0$. Let θ_0 be the true value of θ.

Theorem 1. *Let $Y_1, Y_2, \ldots$ be a sequence of random variables and $U_i = U(Y_i,\theta_0)$ such that $E(U_i) = 0$, $V(U_i) = v_i\sigma^2$ and $v_i^{-1/2}U_i$ are independent and identically distributed. If $E(U_i^4) < \infty$, $\sum_{i=1}^n v_i^2/(\sum_{i=1}^n v_i)^2 = o(1)$ and $n^{-1}\sum_{i=1}^n v_i^{-2} = O(1)$, then $-2\{l_W(\theta_0) - l_W(\hat{\theta})\}$ converges to χ_1^2 in distribution as $n \to \infty$.*

Proof. See the appendix.

The assumptions that the $v_i^{-1/2}U_i$ are iid, $E(U_i^4) < \infty$, and $\sum_{i=1}^n v_i^2/(\sum_{i=1}^n v_i)^2 = o(1)$ can altogether be replaced by a version of Lindeberg's condition. Such a condition, however, will generally be difficult to verify. The condition $\sum_{i=1}^n v_i^2/(\sum_{i=1}^n v_i)^2 = o(1)$ can be sufficiently replaced by $n^{-1}\sum_{i=1}^n v_i^2 = O(1)$. The theorem is most useful under the context of regression analysis where the U_i are often related to the error terms. For instance, the common mean $\mu_0 = E(Y_i)$ can be related to the regression model $Y_i = \mu_0 + v_i^{1/2}\varepsilon_i$. The conditions required by the theorem simply state that the error terms ε_i are iid with finite forth moment.

A unique solution to the constrained maximization problem exists if $U_{(1)} < 0 < U_{(n)}$, where $U_{(1)} = \min\{U_1,\ldots,U_n\}$ and $U_{(n)} = \max\{U_1,\ldots,U_n\}$. This occurs with probability approaching to 1 as $n \to \infty$. It is also clear from the proof of the theorem that one may choose $C_n = (n-1)/\sum_{i=1}^n v_i$ as the scaling constant to match the unbiased estimator for σ^2. This C_n was used in the simulation study reported in Section 3.

The major computational task is to maximize $l_W(F)$ under the constraints $\sum_{i=1}^n p_i = 1$ ($p_i > 0$) and $\sum_{i=1}^n p_i U(Y_i,\theta) = 0$. It can easily be seen through the Lagrange multiplier method that, unlike the usual empirical likelihood approach, the Lagrange multiplier corresponding to the normalization constraint $\sum_{i=1}^n p_i = 1$ cannot be eliminated under the current context. We need to combine $\sum_{i=1}^n p_i = 1$ and $\sum_{i=1}^n p_i U(Y_i,\theta) = 0$ together to form a single set of constraints as $\sum_{i=1}^n p_i \boldsymbol{z}_i = \boldsymbol{Z}$, where $\boldsymbol{z}_i = (1, U_i)'$ and $\boldsymbol{Z} = (1,0)'$. It can be shown by using the Lagrange multiplier method that the p_i which maximize $l_W(F)$ subject to $\sum_{i=1}^n p_i \boldsymbol{z}_i = \boldsymbol{Z}$ are given by $p_i = \{n(1 + \boldsymbol{\lambda}'\boldsymbol{z}_i q_i)\}^{-1}$, where $q_i = v_i^{-1}$, and the Lagrange multiplier $\boldsymbol{\lambda}$ is the solution to

$$g(\boldsymbol{\lambda}) = \frac{1}{n}\sum_{i=1}^n \frac{\boldsymbol{z}_i}{1 + \boldsymbol{\lambda}'\boldsymbol{z}_i q_i} - \boldsymbol{Z} = \boldsymbol{0}. \tag{2.1}$$

It should be noted that the convex duality property from the usual empirical likelihood method is no longer true for the weighted empirical likelihood approach proposed here. A modified Newton-Raphson algorithm similar to the one introduced in Wu (2003a), however, can be used for solving (2.1). Note that the constraints $p_i > 0$ and $\sum_{i=1}^n p_i = 1$ require that $1 + \boldsymbol{\lambda}'\boldsymbol{z}_i q_i > n^{-1}$ for all i. This is a crucial requirement that should be checked at each updating step when the conventional Newton–Raphson method is used to solve $g(\boldsymbol{\lambda}) = \boldsymbol{0}$.

Step 0: Let $\boldsymbol{\lambda}_0 = \mathbf{0}$. Set $\varepsilon = 10^{-8}$.

Step 1: Calculate $\Delta(\boldsymbol{\lambda}_k)$ where

$$\Delta(\boldsymbol{\lambda}) = \left\{ -\frac{1}{n}\sum_{i=1}^{n} \frac{q_i \boldsymbol{z}_i \boldsymbol{z}_i'}{(1+\boldsymbol{\lambda}'\boldsymbol{z}_i q_i)^2} \right\}^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} \frac{\boldsymbol{z}_i}{1+\boldsymbol{\lambda}'\boldsymbol{z}_i q_i} - \boldsymbol{Z} \right).$$

If $\|\Delta(\boldsymbol{\lambda}_k)\| < \varepsilon$, stop the algorithm and report $\boldsymbol{\lambda}_k$; otherwise go to Step 2.

Step 2: Let $\boldsymbol{\delta}_k = \Delta(\boldsymbol{\lambda}_k)$. If $1 + (\boldsymbol{\lambda}_k - \boldsymbol{\delta}_k)'\boldsymbol{z}_i q_i \leqslant n^{-1}$ for some i, let $\boldsymbol{\delta}_k = \boldsymbol{\delta}_k/2$ and repeat Step 2.

Step 3: Set $\boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k - \boldsymbol{\delta}_k$ and $k = k+1$. Go to Step 1.

Our experience from the reported simulation studies shows that this procedure works very well, and is indeed very efficient when $\boldsymbol{Z}$ is not near the boundary of the convex hull formed by $\{\boldsymbol{z}_i : 1 \leqslant i \leqslant n\}$.

The weighted empirical likelihood ratio confidence interval for θ_0 in the form of

$$\{\theta | -2[l_{\mathrm{W}}(\theta) - l_{\mathrm{W}}(\hat{\theta})] < \chi_1^2(\alpha)\} \tag{2.2}$$

is now justified to have approximate coverage probability of $1-\alpha$ for large samples. The two confidence intervals, the conventional unweighted one and (2.2), however, do behave differently when sample size is small. We will illustrate this in Section 3 using a simple linear regression model where the weighted and the unweighted empirical likelihood ratio confidence intervals for the regression coefficient β are examined through a simulation study.

3. Confidence intervals for a regression coefficient

In this section we consider a simple linear regression model $Y = \beta x + v^{1/2}\varepsilon$ where $E(\varepsilon \,|\, x) = 0$, $V(\varepsilon \,|\, x) = \sigma^2$. Typically $v = v(x)$ for some known function $v(\cdot)$, and the exact distribution of ε is left unspecified. We are interested in constructing a $1-\alpha$ confidence interval for β.

Let $\{(Y_i, x_i),\ i=1,\ldots,n\}$ be an independent sample from the regression model. Let $U_i(\beta) = Y_i - \beta x_i$, $i = 1,\ldots,n$. It follows that $E\{U_i(\beta)\} = 0$ and $V\{U_i(\beta)\} = v_i\sigma^2$, where β is the true value of the regression coefficient. The weighted empirical likelihood ratio confidence interval for β, denoted by WL, can be constructed in the form of (2.2). The usual unweighted empirical likelihood ratio confidence interval for β is denoted by EL, where the normal equation $\sum_{i=1}^{n} x_i(Y_i - \beta x_i)/v_i = 0$ is used as constraint. A normal confidence interval based on the weighted least-squares estimator and its estimated variance for β is given by

$$\{\hat{\beta} - Z_{\alpha/2}\,\mathrm{SE}(\hat{\beta}), \quad \hat{\beta} + Z_{\alpha/2}\,\mathrm{SE}(\hat{\beta})\}, \tag{3.1}$$

where $\hat{\beta} = (\sum_{i=1}^{n} x_i Y_i/v_i)/(\sum_{i=1}^{n} x_i^2/v_i)$, and $\mathrm{SE}(\hat{\beta})$ is the standard error of $\hat{\beta}$ computed using standard weighted least-squares theory. This interval is denoted by NC.

The $1-\alpha$ coverage probability for all three intervals WL, EL and NC is justified for large samples, and therefore the three intervals are all asymptotically valid. When sample size is small or moderate, however, these confidence intervals behave quite differently, as shown from the simulation results reported below.

At each simulation run, a sample of size n is generated from the model $Y_i = \beta x_i + |x_i|^{1/2}\varepsilon_i$ with the true value of β setting to 1. The covariates x_i are generated from a standard gamma distribution. Two distributions are used to generate the error terms ε_i: the symmetric distribution $\mathrm{N}(0,1)$ and the

Table 1
Performance of 90% confidence intervals for β

n	CI	CP	L	U	AL
20	NC	84.8	2.6	12.6	1.03
	EL	81.4	5.1	13.5	0.94
	WL	84.5	5.9	9.6	1.01
40	NC	86.9	2.7	10.4	0.72
	EL	84.8	4.9	10.3	0.70
	WL	87.5	6.1	6.4	0.75
100	NC	88.8	3.2	8.0	0.46
	EL	88.2	4.5	7.3	0.46
	WL	89.9	5.8	4.3	0.48

right skewed distribution $\chi_1^2 - 1$. Let (L_b, U_b) be a confidence interval (CI) computed from the bth simulated sample for $b = 1, \ldots, B$. The total number of simulation runs is $B = 5000$. The simulated coverage probability (CP), lower side tail error (L), upper side tail error (U), average length of the interval (AL) are computed as $\mathrm{CP} = B^{-1} \sum_{b=1}^{B} I(L_b < \beta < U_b)$, $L = B^{-1} \sum_{b=1}^{B} I(\beta \leqslant L_b)$, $U = B^{-1} \sum_{b=1}^{B} I(\beta \geqslant U_b)$, and $\mathrm{AL} = B^{-1} \sum_{b=1}^{B} (U_b - L_b)$, respectively. Note that $\mathrm{CP} + L + U = 1$. The simulated results of 90% confidence intervals for β with $\varepsilon_i \sim \chi_1^2 - 1$ under various sample size n are reported in Table 1.

Table 1 can be summarized as follows: (i) the coverage probability for the conventional EL interval is lower than the nominal value 90% at all cases and is as low as 81.4% for $n = 20$. The associated two-sided tail errors are also not balanced; (ii) the normal confidence interval NC has better performance than that of the EL in terms of coverage probability but the two-sided tail errors are severely unbalanced; (iii) the weighted WL interval has the best coverage probability among the three intervals and is also most balanced in terms of tail errors; (iv) the better coverage probability of the WL interval is partially due to the fact that it is slightly wider than the EL interval. The balanced tail behavior of the WL interval, however, remains unexplained.

As for the case of $\varepsilon_i \sim \mathrm{N}(0, 1)$ for which the results are not reported here, the WL interval performs similar to the NC interval and both intervals perform well. The under-coverage problem for the EL interval is still evident when $n \leqslant 60$, although it is less severe than those seen in Table 1. The two-sided tail errors are quite balanced for all three intervals.

There exist more sophisticated approaches that can be combined to the usual empirical likelihood ratio confidence interval to improve the coverage probabilities. For example, several bootstrap related methods as described in Owen (1988) or a second order adjustment such as the Bartlett correction can be used to achieve this goal. The weighted empirical likelihood approach proposed here can be viewed as a special adjustment for the heteroscedastic variance structure of the data. It has the major advantage of computational simplicity. It should be noted, however, that our proposed approach is different from those used by Kolaczyk (1994) and Chen and Cui (2003) where weighting is applied to constraints, not directly to the empirical likelihood function. Our approach is also very useful and generally applicable for efficient point estimation of a population mean when side information is available. This is detailed in the next section.

4. Estimation of a population mean using side information

It turns out that the weighted empirical likelihood method provides a very useful alternative approach for point estimation of a population mean when side information is available. This is of particular interest in the context of survey sampling where auxiliary information is routinely used through the so-called calibration method. Chen and Qin (1993), Chen and Sitter (1999) and Zhong and Rao (2000) contain detailed treatment of the empirical likelihood method in survey sampling. See also Wu (2003a) for a discussion. In this section we show that the proposed weighted maximum empirical likelihood estimator defined shortly provides considerable gain in efficiency over the unweighted estimator when the sample size is not large.

4.1. Infinite populations

Suppose Y is the variable of interest with unknown mean μ_y, and $\boldsymbol{x}$ is a vector of covariates with known mean values $\boldsymbol{\mu}_x$. Let $\{(y_i, \boldsymbol{x}_i),\ i = 1, \ldots, n\}$ be a random sample. The inference problem here is to improve the point estimate of μ_y using the known side information $\boldsymbol{\mu}_x$. Such a scenario is most commonly seen in survey sampling where the covariates $\boldsymbol{x}$ are also termed auxiliary variables, and the population means $\boldsymbol{\mu}_x$ are often available from census or other sources.

The empirical likelihood method provides a natural way of incorporating $\boldsymbol{\mu}_x$ for the estimation of μ_y through constrained maximum likelihood estimation. The unweighted maximum empirical likelihood estimator of μ_y is computed as $\hat{\mu}_y = \sum_{i=1}^n p_i y_i$ where the p_i maximize the empirical likelihood function $l(F) = \sum_{i=1}^n \log(p_i)$ subject to

$$\sum_{i=1}^n p_i = 1 \quad (p_i > 0) \quad \text{and} \quad \sum_{i=1}^n p_i \boldsymbol{x}_i = \boldsymbol{\mu}_x. \tag{4.1}$$

The two constraints in (4.1) may be combined together and the computation may be carried through as in the previous section. We can also treat the unweighted empirical likelihood method as a special case of the weighted empirical likelihood approach using $v_i = 1$. From the proof of Theorem 2 in the appendix we see that

$$\hat{\mu}_y = \bar{y} + \hat{\boldsymbol{\beta}}'(\boldsymbol{\mu}_x - \bar{\boldsymbol{x}}) + \mathrm{O_p}(n^{-1}), \tag{4.2}$$

where $\bar{y}$ and $\bar{\boldsymbol{x}}$ are the sample means, and $\hat{\boldsymbol{\beta}} = (\sum_{i=1}^n \boldsymbol{x}_i \boldsymbol{x}_i')^{-1} \sum_{i=1}^n \boldsymbol{x}_i y_i$ is the ordinary least-squares estimator for the regression coefficients $\boldsymbol{\beta}$ associated with the underlying model

$$y_i = \boldsymbol{x}_i' \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \ldots, n, \tag{4.3}$$

where the ε_i are independent given the $\boldsymbol{x}_i$ and $E(\varepsilon_i \mid \boldsymbol{x}_i) = 0$.

Suppose the regression model (4.3) has a heteroscedastic variance structure, i.e. $V(\varepsilon_i \mid \boldsymbol{x}_i) = v_i \sigma^2$. The unweighted estimator $\hat{\boldsymbol{\beta}}$ is no longer optimal under this model. One is compelled to replacing $\hat{\boldsymbol{\beta}}$ by the (weighted) best linear unbiased estimator (BLUE) $\hat{\boldsymbol{\beta}}_{\mathrm{W}} = (\sum_{i=1}^n v_i^{-1} \boldsymbol{x}_i \boldsymbol{x}_i')^{-1} (\sum_{i=1}^n v_i^{-1} \boldsymbol{x}_i y_i)$ in the formulation of $\hat{\mu}_y$ if such a replacement can be done within the same framework of maximum empirical likelihood estimation. This is indeed one of the motivations behind the weighted empirical

likelihood method. If $\tilde{p}_i$ maximize $l_{\mathrm{W}}(F) = C_n \sum_{i=1}^n v_i\{\log(p_i) - n p_i\}$ subject to the same set of constraints (4.1), the resulting weighted maximum empirical likelihood estimator of μ_y is given by $\tilde{\mu}_y = \sum_{i=1}^n \tilde{p}_i y_i = \bar{y} + \hat{\boldsymbol{\beta}}_{\mathrm{W}}'(\boldsymbol{\mu}_x - \bar{\boldsymbol{x}}) + \mathrm{O_p}(n^{-1})$. See the proof of Theorem 2 in the appendix for arguments leading to this.

The aforementioned expansions for $\hat{\mu}_y$ and $\tilde{\mu}_y$ do not constitute a precise comparison between the two estimators. Both estimators are approximately unbiased for μ_y, and their exact variances are not tractable for a fixed sample size n. The two estimators $\hat{\mu}_y$ and $\tilde{\mu}_y$ become indistinguishable as $n \to \infty$.

Theorem 2. *Suppose that* $n^{-1} \sum_{i=1}^n \|\boldsymbol{x}_i\|^6 = \mathrm{O}(1)$ *and* $n^{-1} \sum_{i=1}^n v_i^{-4} = \mathrm{O}(1)$, *then the biases for both* $\hat{\mu}_y$ *and* $\tilde{\mu}_y$ *are of order* $\mathrm{O}(n^{-1})$, *and the two estimators have the same variance up to the order of* $\mathrm{O}(n^{-3/2})$.

Proof. See the appendix.

It follows that the two estimators have the same mean square error up to the high order of $\mathrm{O}(n^{-3/2})$. To weight or not to weight is not a critical issue for large samples. The use of the weighted empirical likelihood estimator $\tilde{\mu}_y$ over the unweighted one $\hat{\mu}_y$, however, has the implicit effect of replacing $\hat{\boldsymbol{\beta}}$ by the best linear unbiased estimator $\hat{\boldsymbol{\beta}}_{\mathrm{W}}$ which is indeed beneficial for the estimation of μ_y when the sample size is not large, as supported by results from a limited simulation study reported below.

Simulated samples $\{(y_i, x_i),\ i = 1, \ldots, n\}$ are generated from the same regression model $Y_i = \beta x_i + |x_i|^{1/2}\varepsilon_i$ used in Section 3. The unweighted and the weighted estimators $\hat{\mu}_y$ and $\tilde{\mu}_y$ are computed based on each of the simulated samples, treating $\mu_x = 1$ as known. The process is independently repeated for $B = 5000$ times. Performance of $\hat{\mu}_y$ (also similarly defined for $\tilde{\mu}_y$) is measured in terms of relative bias (RB) and relative efficiency (RE) defined as

$$\mathrm{RB} = \frac{1}{B}\sum_{b=1}^{B} \frac{\hat{\mu}_y(b) - \mu_y}{\mu_y} \quad \text{and} \quad \mathrm{RE} = \frac{\mathrm{MSE}(\bar{y})}{\mathrm{MSE}(\hat{\mu}_y)},$$

where the true value of μ_y is known under the simulation setting, $\hat{\mu}_y(b)$ is the value of $\hat{\mu}_y$ computed from the bth simulated sample, and $\mathrm{MSE}(\hat{\mu}_y) = B^{-1}\sum_{b=1}^{B}\{\hat{\mu}_y(b) - \mu_y\}^2$ (and $\mathrm{MSE}(\bar{y})$ similarly defined). The sample mean $\bar{y}$ is used as baseline estimator for comparison. Large values of RE (> 1) indicate high efficiency of the estimator compared to $\bar{y}$.

The simulated absolute values of RB are less than 0.7% for all cases and thus are not reported here. Part of Table 2 summarizes the simulated values of RE for various sample sizes and for each of the two error distributions used in the simulation.

The simulation results show that the use of side information (the known μ_x) makes the empirical likelihood estimators $\hat{\mu}_y$ and $\tilde{\mu}_y$ much more efficient than the naive estimator $\bar{y}$. It is also shown clearly that the weighted estimator $\tilde{\mu}_y$ performs uniformly better than the unweighted estimator $\hat{\mu}_y$, with larger gain in efficiency occurring at smaller sample size. The reduction in terms of mean square error (MSE) from using $\tilde{\mu}_y$ over $\hat{\mu}_y$ ranges from about 10% for $n = 20$ to 2.5% for $n = 100$.

Table 2
Simulated relative efficiencies for $\hat{\mu}_y$ and $\tilde{\mu}_y$

		Infinite population			Finite population		
	$n \to$	20	40	100	20	40	100
$\varepsilon \sim \mathrm{N}(0,1)$	$\hat{\mu}_y$	1.64	1.88	1.93	1.60	1.84	1.99
	$\tilde{\mu}_y$	1.78	1.96	1.98	1.72	1.92	2.03
$\varepsilon \sim \chi^2_1 - 1$	$\hat{\mu}_y$	1.20	1.34	1.42	1.22	1.35	1.47
	$\tilde{\mu}_y$	1.33	1.42	1.47	1.34	1.43	1.51

4.2. Finite populations

The asymptotic arguments presented in Theorem 2 do not apply directly to finite populations. The so-called design-based approach in survey sampling, where randomization is induced by repeated sampling from the fixed finite population, imposes certain restrictions for large sample comparisons.

Let $\{(y_i, \boldsymbol{x}_i), i = 1, \ldots, N\}$ be the values of the study variable Y and the vector of auxiliary variables $\boldsymbol{x}$ for the finite population of size N. With some misuse of notation but without causing any confusion, let $\mu_y = N^{-1} \sum_{i=1}^N y_i$ and $\boldsymbol{\mu}_x = N^{-1} \sum_{i=1}^N \boldsymbol{x}_i$ be the finite population means, $\boldsymbol{\beta} = (\sum_{i=1}^N \boldsymbol{x}_i \boldsymbol{x}_i')^{-1} \sum_{i=1}^N \boldsymbol{x}_i y_i$ and $\boldsymbol{\beta}_{\mathrm{W}} = (\sum_{i=1}^N \boldsymbol{x}_i \boldsymbol{x}_i'/v_i)^{-1} \sum_{i=1}^N \boldsymbol{x}_i y_i/v_i$ be the unweighted and the weighted finite population regression coefficients. Our goal is to estimate μ_y using survey sample data $\{(y_i, \boldsymbol{x}_i), i \in s\}$, where s is the set of n sampled elements from the finite population. The mean values $\boldsymbol{\mu}_x$ are assumed known. We restrict our discussion to cases where s is drawn by simple random sampling without replacement. Wu (2003a) contains a brief discussion on the formulation of the weighted empirical likelihood method in survey sampling under an arbitrary probability sampling design.

The unweighted and the weighted maximum empirical likelihood estimators $\hat{\mu}_y$ and $\tilde{\mu}_y$ are once again computed as $\sum_{i=1}^n p_i y_i$, where the p_i maximize $l(F)$ and $l_{\mathrm{W}}(F)$, respectively, subject to the set of constraints (4.1).

Our discussion below requires suitable asymptotic set-up that allows $n \to \infty$ under the framework of finite populations. We refer to Isaki and Fuller (1982) for a detailed formulation. Under the same finite moment conditions used in Theorem 2, we have

$$\hat{\mu}_y = \bar{y} + \hat{\boldsymbol{\beta}}'(\boldsymbol{\mu}_{\boldsymbol{x}} - \bar{\boldsymbol{x}}) + \mathrm{O_p}(n^{-1}) = \bar{y} + \boldsymbol{\beta}'(\boldsymbol{\mu}_{\boldsymbol{x}} - \bar{\boldsymbol{x}}) + \mathrm{O_p}(n^{-1}),$$

$$\tilde{\mu}_y = \bar{y} + \hat{\boldsymbol{\beta}}_{\mathrm{W}}'(\boldsymbol{\mu}_{\boldsymbol{x}} - \bar{\boldsymbol{x}}) + \mathrm{O_p}(n^{-1}) = \bar{y} + \boldsymbol{\beta}_{\mathrm{W}}'(\boldsymbol{\mu}_{\boldsymbol{x}} - \bar{\boldsymbol{x}}) + \mathrm{O_p}(n^{-1}).$$

The stochastic order $\mathrm{O_p}(\cdot)$ refers to the probability sampling under the design-based approach.

It is now evident that the two estimators $\hat{\mu}_y$ and $\hat{\mu}_y$ have the same order of bias at $\mathrm{O}(n^{-1})$ but their design-based variances are different even at the first order of $\mathrm{O}(n^{-1})$. More specifically, let $V_p(\cdot)$ denote the design-based variance, we have

$$V_p(\hat{\mu}_y) = (1 - f)n^{-1} V_e + \mathrm{O}(n^{-3/2}) \quad \text{and} \quad V_p(\tilde{\mu}_y) = (1 - f)n^{-1} V_r + \mathrm{O}(n^{-3/2}),$$

where $V_e=(N-1)^{-1}\sum_{i=1}^{N}(e_i-\bar{e}_N)^2$, $V_r=(N-1)^{-1}\sum_{i=1}^{N}(r_i-\bar{r}_N)^2$, $e_i=y_i-\boldsymbol{\beta}'\boldsymbol{x}_i$, $\bar{e}_N=N^{-1}\sum_{i=1}^{N}e_i$, $r_i=y_i-\boldsymbol{\beta}_{\mathrm{W}}'\boldsymbol{x}_i$, $\bar{r}_N=N^{-1}\sum_{i=1}^{N}r_i$, and $1-f=1-n/N$ is the finite population correction factor.

A first-order comparison between $V_p(\hat{\mu}_y)$ and $V_p(\tilde{\mu}_y)$ under the design-based approach is not feasible, since the difference lies between the two versions of regression coefficients $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_{\mathrm{W}}$ in defining V_e and V_r, and such a difference can only be assessed under the regression model (4.3). A model such as (4.3) is also termed as superpopulation model in the survey sampling context. The finite population is viewed as independent realizations from the superpopulation model. Large sample comparisons, however, can often be made using the so-called anticipated variances (Isaki and Fuller, 1982) $E_\xi\{V_p(\hat{\mu}_y)\}$ and $E_\xi\{V_p(\tilde{\mu}_y)\}$, where $E_\xi(\cdot)$ represents the conditional expectation under the superpopulation model given the $\boldsymbol{x}_i$, $i=1,\ldots,N$. The anticipated variance $E_\xi\{V_p(\cdot)\}$ is the design-based variance for a particular finite population, averaged over all possible realizations of such populations under the superpopulation model.

Theorem 3. *Suppose that* $N^{-1}\sum_{i=1}^{N}\|\boldsymbol{x}_i\|^4=\mathrm{O}(1)$, $N^{-1}\sum_{i=1}^{N}v_i=\mathrm{O}(1)$, $N^{-1}\sum_{i=1}^{N}v_i^{-4}=\mathrm{O}(1)$, *then* $E_\xi(V_e)=E_\xi(V_r)+\mathrm{O}(N^{-1})$, *where* ξ *denotes the regression model* (4.3).

Proof. The proof involves some lengthy matrix manipulation and can be found in Wu (2003b).

Hence the two estimators have virtually the same anticipated variance. The first-order difference between their design-based variances may not have substantial consequence even under the design-based framework if the sample size is large.

To further explore their small sample design-based performances, we modify the simulation study of Section 4.1 as follows. First, a finite population of size $N=2000$ is generated from the same regression model used in Section 4.1, and this population is treated as fixed. The true values of μ_y and μ_x are determined from this particular population. Repeated samples are then drawn from this fixed population by simple random sampling without replacement, and the estimators $\tilde{\mu}_y$ and $\hat{\mu}_y$ are computed for each of the simulated samples. The total number of repeated samples is $B=5000$. The absolute values of the simulated relative biases are all less than 1% and in most cases are smaller than 0.1%. The relative efficiencies are reported in Table 2. The message conveyed here resembles those of Section 4.1: the weighted estimator $\tilde{\mu}_y$ performs uniformly better than the unweighted one $\hat{\mu}_y$, and the gain in efficiency is higher when the sample size is smaller.

5. Additional remarks

Weighting is commonly used in statistics to account for specific structure of the data. The weighted empirical likelihood approach proposed in this article can be adapted to other type of likelihood function such as the Euclidean likelihood briefly mentioned in Section 2, and the associated large sample properties can similarly be established. The empirical likelihood function is often preferred due to its similarity to the parametric likelihood, and the natural constraints $p_i>0$ and $\sum_{i=1}^{n}p_i=1$. The latter is particularly attractive for point estimation. Obtaining normalized positive weights is a constant theme for estimation in survey sampling. While such constraints can also be imposed to any other approach, the empirical likelihood method achieves these with simple algorithms. The modified Newton–Raphson algorithm described in Section 2 can easily be programmed using statistical

softwares such as SAS or R/Splus. The simulation studies reported in this article are programmed in R/Splus, and the source codes are available from the author.

There is strong evidence in favor of a weighted approach when the sample size is not large and the heteroscedastic structure of the data can clearly be identified. More work is needed, however, to explore the robustness of the approach under misspecified variance structure, and to extend the idea of weighting to more complex situations.

Acknowledgements

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada. Stimulating discussions with Professor Jiahua Chen and helpful comments from a referee are also delightfully acknowledged.

Appendix. Proofs

Proof of Theorem 1. We assume $U_{(1)} < 0 < U_{(n)}$ so that the maximum weighted empirical likelihood solution exists. There are two major aspects involved in our proof which are different from the usual empirical likelihood method. First, as mentioned in Section 2, the Lagrange multiplier corresponding to $\sum_{i=1}^n p_i = 1$ cannot be eliminated under the current context; and secondly, the weighting factor v_i needs to be treated with care at various points of the proof.

If we substitute $\boldsymbol{z}_i/(1+\boldsymbol{\lambda}'\boldsymbol{z}_i q_i)$ by $\boldsymbol{z}_i - q_i\boldsymbol{z}_i\boldsymbol{z}_i'\boldsymbol{\lambda}/(1+\boldsymbol{\lambda}'\boldsymbol{z}_i q_i)$, we can re-write (2.1) as $\tilde{S}\boldsymbol{\lambda} = \bar{\boldsymbol{z}} - \boldsymbol{Z}$, where $\bar{\boldsymbol{z}} = n^{-1}\sum_{i=1}^n \boldsymbol{z}_i$ and

$$\tilde{S} = \frac{1}{n}\sum_{i=1}^n \frac{q_i\boldsymbol{z}_i\boldsymbol{z}_i'}{1+\boldsymbol{\lambda}'\boldsymbol{z}_i q_i}.$$

Let $B_n^2 = \sum_{i=1}^n V(U_i) = \sigma^2\sum_{i=1}^n v_i$. The assumptions $\sum_{i=1}^n v_i^2/(\sum_{i=1}^n v_i)^2 = \mathrm{o}(1)$ and $E(U_i^4) < \infty$ imply that

$$\sum_{i=1}^n E(U_i^4) = \sigma_4\sum_{i=1}^n v_i^2 = \mathrm{o}(B_n^4),$$

where $\sigma_4 = E\{(v_i^{-1/2}U_i)^4\}$ is finite and independent of i. So the Liapunov's condition holds for U_i, which in turn implies the Lindeberg's condition for U_i. By the central limit theorem, $\bar{U} = n^{-1}\sum_{i=1}^n U_i = \mathrm{O_p}(n^{-1/2})$, and hence $\bar{\boldsymbol{z}} - \boldsymbol{Z} = (0, \bar{U})' = \mathrm{O_p}(n^{-1/2})$. The conditions $E(U_i^4) < \infty$ and $n^{-1}\sum_{i=1}^n v_i^{-2} = \mathrm{O}(1)$ also imply that $\max_{1\leqslant i\leqslant n}\|q_i\boldsymbol{z}_i\| = \mathrm{o_p}(n^{1/2})$. Following similar arguments as in Owen (2001, p. 220) we can show that $\|\boldsymbol{\lambda}\| = \mathrm{O_p}(n^{-1/2})$ and $\boldsymbol{\lambda} = S^{-1}(\bar{\boldsymbol{z}} - \boldsymbol{Z}) + \mathrm{o_p}(n^{-1/2})$, where $S = n^{-1}\sum_{i=1}^n q_i\boldsymbol{z}_i\boldsymbol{z}_i'$. We also conclude that $\max_{1\leqslant i\leqslant n}|\boldsymbol{\lambda}'\boldsymbol{z}_i q_i| = \mathrm{o_p}(1)$ under the given conditions.

To finish the proof, we note that $\log(1+u_i) = u_i - u_i^2/2 + \eta_i$ and $(1+u_i)^{-1} = 1 - u_i + u_i^2 + \gamma_i$ if $u_i = \mathrm{o_p}(1)$, where the remainder terms η_i and γ_i can be treated similarly as in Owen (2001, p. 221),

we have

$$-2\{l_{\mathrm{W}}(\theta_0) - l_{\mathrm{W}}(\hat{\theta})\} = 2C_n \sum_{i=1}^{n} q_i^{-1}\{\log(1 + \boldsymbol{\lambda}'\boldsymbol{z}_i q_i) + (1 + \boldsymbol{\lambda}'\boldsymbol{z}_i q_i)^{-1} - 1\}$$

$$= C_n \left\{ \boldsymbol{\lambda}' \left(\sum_{i=1}^{n} q_i \boldsymbol{z}_i \boldsymbol{z}_i' \right) \boldsymbol{\lambda} \right\} + \mathrm{o_p}(1)$$

$$= nC_n(\bar{\boldsymbol{z}} - \boldsymbol{Z})' S^{-1}(\bar{\boldsymbol{z}} - \boldsymbol{Z}) + \mathrm{o_p}(1)$$

$$= (\bar{U})^2/B + \mathrm{o_p}(1),$$

where $B = n^{-3} \sum_{i=1}^{n} v_i \sum_{i=1}^{n} q_i(U_i - \bar{U}_{\mathrm{W}})^2$, $\bar{U}_{\mathrm{W}} = \sum_{i=1}^{n} q_i U_i / \sum_{i=1}^{n} q_i$. The very last step follows by noting that $\bar{\boldsymbol{z}} - \boldsymbol{Z} = (0, \bar{U})'$, and the lower right corner element in the two by two matrix S^{-1} is $\{n^{-1} \sum_{i=1}^{n} q_i(U_i - \bar{U}_{\mathrm{W}})^2\}^{-1}$. The final conclusion that $(\bar{U})^2/B$ converges in distribution to χ_1^2 follows from the fact that $E(\bar{U}) = 0$, $V(\bar{U}) = n^{-2} \sum_{i=1}^{n} v_i \sigma^2$, $E\{\sum_{i=1}^{n} q_i(U_i - \bar{U}_{\mathrm{W}})^2\} = (n-1)\sigma^2$ and applying the central limit theorem to $\bar{U}$. □

Proof of Theorem 2. The key argument is to show that the Lagrange multiplier $\boldsymbol{\lambda}$ can be approximated by $S^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_x)$ with an error of order $\mathrm{O_p}(n^{-1})$ (the commonly claimed order is $\mathrm{o_p}(n^{-1/2})$). We show this for the case of weighted empirical likelihood method. The unweighted case amounts to setting $v_i = 1$. Let $q_i = v_i^{-1}$.

Let $\boldsymbol{x}_i$ be augmented to include 1 as its first component. The weighted empirical likelihood estimator of μ_y is computed as $\tilde{\mu}_y = \sum_{i=1}^{n} p_i y_i$ where $p_i = \{n(1 + \boldsymbol{\lambda}'\boldsymbol{x}_i q_i)\}^{-1}$ with $\boldsymbol{\lambda}$ satisfying (A): $\boldsymbol{\mu}_x = n^{-1} \sum_{i=1}^{n} (1 + \boldsymbol{\lambda}'\boldsymbol{x}_i q_i)^{-1} \boldsymbol{x}_i$. Apply the identity $(1 + \boldsymbol{\lambda}'\boldsymbol{x}_i q_i)^{-1} = 1 - (1 + \boldsymbol{\lambda}'\boldsymbol{x}_i q_i)^{-1} q_i \boldsymbol{x}_i' \boldsymbol{\lambda}$ twice to (A) we get

$$\bar{\boldsymbol{x}} - \boldsymbol{\mu}_x = S\boldsymbol{\lambda} - \|\boldsymbol{\lambda}\|^2 \frac{1}{n} \sum_{i=1}^{n} (1 + \boldsymbol{\lambda}'\boldsymbol{x}_i q_i)^{-1} q_i^2 (\boldsymbol{x}_i'\phi) \boldsymbol{x}_i \boldsymbol{x}_i' \phi,$$

where $S = n^{-1} \sum_{i=1}^{n} q_i \boldsymbol{x}_i \boldsymbol{x}_i'$ and $\boldsymbol{\lambda} = \|\boldsymbol{\lambda}\|\phi$ for some unit vector ϕ. Note that $(1 + \boldsymbol{\lambda}'\boldsymbol{x}_i q_i)^{-1} = 1 + \mathrm{o_p}(1)$, with the term $\mathrm{o_p}(1)$ uniformly over $1 \leqslant i \leqslant n$. Under the conditions of the theorem we have $n^{-1} \sum_{i=1}^{n} q_i^2 \|\boldsymbol{x}_i\|^3 = \mathrm{O}(1)$, it follows from $\|\boldsymbol{\lambda}\| = \mathrm{O}(n^{-1/2})$ and $S = \mathrm{O}(1)$ that $\boldsymbol{\lambda} = S^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_x) + \mathrm{O_p}(n^{-1})$.

Note that, if A and B satisfy $V(A) = \mathrm{O}(n^{-1})$ and $B = \mathrm{O_p}(n^{-1})$, then $V(A+B) = V(A) + \mathrm{O}(n^{-3/2})$. It is a straightforward expansion to show that $\tilde{\mu}_y = \bar{y} + \hat{\boldsymbol{\beta}}_{\mathrm{W}}'(\boldsymbol{\mu}_x - \bar{\boldsymbol{x}}) + \mathrm{O_p}(n^{-1}) = \bar{y} + \boldsymbol{\beta}'(\boldsymbol{\mu}_x - \bar{\boldsymbol{x}}) + \mathrm{O_p}(n^{-1})$, and consequently $E(\tilde{\mu}_y) = \mu_y + \mathrm{O}(n^{-1})$ and $V(\tilde{\mu}_y) = V\{\bar{y} + \boldsymbol{\beta}'(\boldsymbol{\mu}_x - \bar{\boldsymbol{x}})\} + \mathrm{O}(n^{-3/2})$. Same arguments hold for $\hat{\mu}_y$. □

References

Chen, S.X., Cui, H., 2003. An extended empirical likelihood for generalized linear models. Statist. Sinica 13, 69–81.

Chen, J., Qin, J., 1993. Empirical likelihood estimation for finite population and the effective usage of auxiliary information. Biometrika 80, 107–116.

Chen, J., Sitter, R.R., 1999. A pseudo empirical likelihood approach to the effective use of auxiliary information in complex surveys. Statist. Sinica 9, 385–406.

Deville, J.C., Särndal, C.E., 1992. Calibration estimators in survey sampling. J. Amer. Statist. Assoc. 87, 376–382.
Isaki, C.T., Fuller, W.A., 1982. Survey design under the regression superpopulation model. J. Amer. Statist. Assoc. 77, 89–96.
Kolaczyk, E.D., 1994. Empirical likelihood for generalized linear models. Statist. Sinica 4, 199–218.
Owen, A.B., 1988. Empirical likelihood ratio confidence intervals for a single functional. Biometrika 75, 237–249.
Owen, A.B., 2001. Empirical Likelihood. Chapman & Hall/CRC.
Wu, C., 2003a. Some algorithmic aspects of the empirical likelihood method in survey sampling. Statist. Sinica, accepted for publication in May 2003.
Wu, C., 2003b. Weighted empirical likelihood inference. Working Paper 2003-01, Department of Statistics and Actuarial Science, University of Waterloo.
Zhong, B., Rao, J.N.K., 2000. Empirical likelihood inference under stratified random sampling using auxiliary population information. Biometrika 87, 929–938.