Biometrika (20xx), **xx**, x, *pp*. 1–13

Printed in Great Britain

Advance Access publication on 31 July 20xx

Modelling complex survey data with population level information: an empirical likelihood approach

BY M. OGUZ-ALPER

Statistics Norway, Postboks 8131 Dep, NO-0033 Oslo, Norway.
melike.oguz.alper@ssb.no

AND Y. G. BERGER

Southampton Statistical Sciences Research Institute,
University of Southampton, SO17 1BJ Southampton, United Kingdom
y.g.berger@soton.ac.uk

SUMMARY

Survey data are often collected with unequal probabilities from a stratified population. In many modelling situations, the parameter of interest is a subset of a set of parameters, with the others treated as nuisance parameters. We show that in this situation the empirical likelihood ratio statistic follows a chi-squared distribution asymptotically, under stratified single and multi-stage unequal probability sampling, with negligible sampling fractions. Simulation studies show that the empirical likelihood confidence interval may achieve better coverages and has more balanced tail error rates than standard approaches, which involve variance estimation, linearization or re-sampling.

Some key words: Design-based inference; empirical likelihood; estimating equation; inclusion probability; regression parameter; unequal probability sampling.

1. INTRODUCTION

Suppose we wish to fit a model to sample data selected randomly with unequal probabilities. When the random selection is ignored, estimators based on the assumption of independent and identically distributed observations may produce invalid inferences, especially if the sampling design is informative (e.g., Pfeffermann & Sverchkov, 1999, 2003).

We consider a design-based inference using empirical likelihood, with parameters defined through population estimating equations. These allow complex parameters such as linear, non-linear or generalized linear regression parameters to be estimated. We derive the asymptotic properties under a sequence of finite populations and a sequence of sampling designs which allow the sample sizes to tend to infinity (Isaki & Fuller, 1982); $o_p(\cdot)$ and $O_p(\cdot)$ denote the orders of convergence in probability with respect to the sampling design.

Let U be a finite population of units labelled $i = 1, \ldots, N$. Suppose that U is stratified into H non-overlapping strata denoted by U_h, where $h = 1, \ldots, H$ and $\cup_{h=1}^{H} U_h = U$. We assume that H is bounded. Suppose that n_h units are selected independently with replacement with unequal probabilities p_i from U_h, where $\sum_{i \in U_h} p_i = 1$. Let $s = \cup_{h=1}^{H} s_h$ denote the sample containing the labels of the units selected after $n = \sum_{h=1}^{H} n_h$ draws. The probability distribution of s is

specified by the sampling design. Let v_i be a vector containing the values of a set of variables for a unit $i \in U$. Under a design-based inference, the v_i are fixed, non-random, vectors.

It is common practice to treat without-replacement samples as if they were selected with replacement, as long as the sampling fraction, $\bar{\pi} = n/N$, is negligible. This implies a negligible bias and slightly more conservative confidence intervals (e.g., Durbin, 1953; Gabler, 1984). Sampling with replacement and large entropy sampling without replacement (Berger, 2011) are equivalent when the sampling fraction is negligible (Hájek, 1981, p.112). Large-entropy designs are mostly used in practice and can be considered as asymptotic designs (Berger, 2011). The approach proposed here is valid under these sampling designs.

Empirical likelihood is well-established under the classical independent and identical distribution framework (Owen, 2001). Several empirical likelihood ratio statistics have been proposed for unequal probability sampling, mostly for single parameters. The pseudo-empirical likelihood ratio statistic (Wu & Rao, 2006) is based on design effects. Kim (2009) and Chen & Kim (2014) developed an approach based on Poisson sampling, which is different from the sampling design considered in the present paper. The generalized pseudo-empirical likelihood ratio statistic (Tan & Wu, 2015) does not need design effects under large entropy sampling. Berger & De La Riva Torres (2016) proposed a different approach, which again does not rely on design effects. The present proposal is based on Berger & De La Riva Torres's (2016) approach and deals with multivariate parameters by profiling.

Profiling consists in minimizing the empirical likelihood ratio statistic over any nuisance parameters in the model, thereby allowing inference for the parameter of interest. Qin & Lawless's (1994) profiling approach is limited to independent and identically distributed observations and does not account for the sampling design and the unequal probabilities. We propose to profile out the empirical likelihood ratio statistic proposed by Berger & De La Riva Torres (2016). We show that the empirical likelihood ratio statistic, evaluated at the true value of the parameter of interest, follows a chi-squared distribution asymptotically.

Binder & Patak (1994) proposed a non-parametric version of likelihood-based score statistics that can be used with nuisance parameters and relies on variance estimates. The bounds of confidence intervals are the solutions to a system of equations that can be solved numerically. Godambe & Thompson (2009, p.92) pointed out that solutions may not exist.

Chen & Sitter (1999) and Zhong & Rao (2000) proposed an algorithm based on profiling the pseudo-empirical likelihood ratio statistic when stratum totals of auxiliary variables are unknown. For confidence intervals, the pseudo-empirical likelihood ratio statistic must be adjusted by variance estimates. The pseudo-empirical likelihood and generalized pseudo-empirical likelihood approaches (Tan & Wu, 2015) are limited to univariate estimating equations. There is no general multivariate theory on profiling for these approaches.

Pfeffermann & Sverchkov (1999, 2003) considered a semi-parametric approach that requires modelling the survey weights. The variances of the model parameters are estimated through linearization or re-sampling (e.g., Pfeffermann & Sverchkov, 1999). In Section 7, we compare our approach with that of Pfeffermann & Sverchkov (1999, 2003).

Standard confidence intervals based on variance estimates may perform poorly, when normality does not hold, as with skewed data and outlying values. Even when normality holds, heteroscedasticity or model misspecification may affect the coverage of standard confidence intervals (Owen, 1991). Furthermore, coverage may be affected by the bias of linearized or re-sampling variance estimators. The proposed confidence interval does not rely on the normality of the point estimator. It is less computer-intensive than the bootstrap and simpler to implement than linearization, because it does not require variance estimation or the derivation of linearized variables. Our simulation studies show that the empirical likelihood confidence interval proposed

has good coverage and balanced tail errors even when the point estimator is not normally distributed.

2. Parameters and estimating equations

The parameter $\psi_N \in \mathbb{R}^b$ is the fixed finite population vector that is the unique solution to the population multivariate estimating equation (Godambe & Thompson, 1986),

$$G(\psi) = \sum_{i \in U} g_i(v_i, \psi) = 0_b, \tag{1}$$

where $g_i(v_i, \psi)$ is a $b \times 1$ finite vector of estimating functions, v_i is the vector of variables for unit i, and 0_b is the $b \times 1$ vector of zeros. We replace $g_i(v_i, \psi)$ by $g_i(\psi)$ for simplicity.

Inferences about the population parameters can be improved by incorporating known population-level parameters, which may be available from administrative data, a census or population projections (e.g., Hartley & Rao, 1969; Owen, 1991; Deville & Särndal, 1992; Chaudhuri et al., 2008). Let φ_N be a vector of known population-level parameters, uniquely defined by the solution to the population estimating equation

$$\sum_{i \in U} f_i(v_i, \varphi) = 0.$$

For simplicity, we replace $f_i(v_i, \varphi)$ by $f_i(\varphi)$. The vector φ_N may be of known population means, totals, ratios, proportions, variances, quantiles or distribution functions of some of the auxiliary variables within v_i; see Berger & De La Riva Torres (2016). We assume that $f_i(\varphi)$ does not depend on ψ_N, and has a finite dimension. The $g_i(\psi)$ may depend on φ_N.

3. Empirical likelihood approach for unequal probability sampling

3·1. *Empirical log-likelihood function*

Consider Berger & De La Riva Torres's (2016) empirical log-likelihood function

$$\ell(m) = \sum_{i \in s} \log m_i, \tag{2}$$

where the m_i are unknown scale-loads allocated to data points $\{i : i \in s\}$ (Hartley & Rao, 1968). Let $\hat{m}_i$ maximize (2) subject to the constraints $m_i > 0$ and

$$\sum_{i \in s} m_i \, c_i = C, \tag{3}$$

with $C = \sum_{i \in U} c_i$ and c_i the r-vector

$$c_i = \left\{ z_i^{\mathrm{T}} \, \bar{\pi}^{\,-1}, f_i(\varphi_N)^{\mathrm{T}} \right\}^{\mathrm{T}}. \tag{4}$$

Here, the H-vector z_i contains the values of the stratification variables; that is,

$$z_i = (z_{i1}, \ldots, z_{iH})^{\mathrm{T}}, \tag{5}$$

where $z_{ih} = \pi_i$ for $i \in U_h$ and $z_{ih} = 0$ otherwise. Here,

$$\pi_i = n_h p_i, \quad \text{when } i \in U_h.$$

We assume that the z_i are known for all the sampled units.

The vector C is known, because $C = (\bar{\pi}^{-1} n_{\mathrm{str}}^{\mathrm{T}}, 0^{\mathrm{T}})^{\mathrm{T}}$, where $n_{\mathrm{str}} = (n_1, \ldots, n_H)^{\mathrm{T}}$ denotes the vector of stratum sample sizes. For the calculation of m_i, $\bar{\pi}^{-1}$ can be removed from c_i and C because it cancels in equation (3). We assume that C is an inner point of the conical hull formed by $\sum_{i \in s} m_i c_i$, so $\hat{m}_i$ is unique. Berger & De La Riva Torres (2016) showed that

$$\hat{m}_i = (\pi_i + \eta^{\mathrm{T}} c_i)^{-1}, \tag{6}$$

where the vector η is such that (3) and $m_i > 0$ hold. A modified Newton–Raphson algorithm (Chen et al., 2002) can be used to compute η.

The $\hat{m}_i$ are always positive and play the role of survey weights. They are also calibrated, because $\sum_{i \in s} \hat{m}_i f_i(\varphi_N) = 0$. The calibration property is the consequence of the maximisation of $\ell(m)$ and the fact that φ_N is known.

The c_i incorporate the information about the sampling design and the population-level parameter. When we use no population-level information and have a single stratum, we have $c_i = \pi_i \bar{\pi}^{-1}$, $\eta = 0$, so $\hat{m}_i = \pi_i^{-1}$ are the basic design weights.

3·2. *Maximum empirical likelihood point estimator*

Let $\hat{m}_i^*(\psi)$ maximize $\ell(m)$ subject to the constraints $m_i > 0$ and

$$\sum_{i \in s} m_i \, c_i^*(\psi) = C^*, \tag{7}$$

with

$$c_i^*(\psi) = \{c_i^{\mathrm{T}}, g_i(\psi)^{\mathrm{T}}\}^{\mathrm{T}}, \quad C^* = (C^{\mathrm{T}}, 0_b^{\mathrm{T}})^{\mathrm{T}}, \tag{8}$$

for a given vector $\psi = (\theta^{\mathrm{T}}, \nu^{\mathrm{T}})^{\mathrm{T}}$; where $g_i(\psi)$ is defined in Section 2. We assume that $c_i^*(\psi)$ is differentiable with respect to ν for all $i \in s$ in a neighbourhood around the true population value ν_N. The maximum value of $\ell(m)$, for a given φ_N, under $m_i > 0$ and the constraint (7) is

$$\ell(\psi, \varphi_N) = \sum_{i \in s} \log \hat{m}_i^*(\psi) \cdot \tag{9}$$

The maximum empirical likelihood estimator $\hat{\psi}$ of ψ_N is the vector that maximizes (9) over ψ. Berger & De La Riva Torres (2016) showed that $\hat{\psi}$ is the solution to

$$\hat{G}(\psi, \varphi_N) = \sum_{i \in s} \hat{m}_i \, g_i(\psi) = 0_b, \tag{10}$$

We assume that $g_i(\psi)$ are such that equation (10) has a unique solution. Specifically, when $\hat{m}_i = \pi_i^{-1}$, the equation (10) is the design-unbiased estimator of $G(\psi)$, for a given ψ, and $\hat{\psi}$ is Binder's (1983) pseudo-likelihood estimator.

4. INFERENCE IN THE PRESENCE OF NUISANCE PARAMETERS

Suppose that we wish to make inference about a $p \times 1$ sub-parameter $\theta_N \in \Theta \subset \mathbb{R}^p$, where $p < b$; that is, $\psi_N = (\theta_N^{\mathrm{T}}, \nu_N^{\mathrm{T}})^{\mathrm{T}}$. The vector ν_N is the $q \times 1$ nuisance parameter, which is not of primary interest. Here, $\nu_N \in \Lambda \subset \mathbb{R}^q$, $q = b - p$; and Θ and Λ are compact sets. The parameter ν_N is assumed unknown and may need to be estimated when making inferences about θ_N. We propose to test and construct a confidence region for the parameter of interest θ_N by using the empirical likelihood ratio statistic,

$$\hat{r}(\theta, \varphi_N) = 2\Big\{\ell(\hat{\psi}, \varphi_N) - \max_{\nu \in \Lambda} \ell(\theta, \nu, \varphi_N)\Big\}, \tag{11}$$

where $\ell(\theta, \nu, \varphi_N) = \ell(\psi, \varphi_N)$ and $\psi = (\theta^{\mathrm{T}}, \nu^{\mathrm{T}})^{\mathrm{T}}$. It can be shown that $\ell(\hat{\psi}, \varphi_N) = \sum_{i \in s} \log \hat{m}_i$, where the $\hat{m}_i$ are defined by (6), because (10) holds for $\hat{\psi}$. In Appendix 1, we propose an algorithm to compute (11).

In Section 5, we show that under regularity conditions and when $\theta = \theta_N$, expression (11) asymptotically follows a chi-squared distribution with p degrees of freedom under unequal probability stratified sampling, where p denotes the dimension of θ_N; that is,

$$\hat{r}(\theta_N, \varphi_N) \to \chi^2_p \tag{12}$$

in distribution, as $n \to \infty$.

The pivotal statistics (11) can also be used to construct confidence intervals for a scalar subparameter θ_N of ψ_N. In this case, ν_N denotes the remaining parameters of ψ_N, so $\hat{r}(\theta_N, \varphi_N)$ follows asymptotically a chi-squared distribution with one degree of freedom. Thus the $\alpha\%$ empirical likelihood confidence interval for θ_N is $\{\theta : \hat{r}(\theta, \varphi_N) \leq \chi^2_1(\alpha)\}$, where $\chi^2_1(\alpha)$ is the upper α-quantile of the chi-squared distribution with one degree of freedom. The $\hat{r}(\theta, \varphi_N)$ is a convex function of θ with a minimum value when $\theta = \hat{\theta}$.

5. Asymptotic properties

5·1. *Regularity conditions*

In this Section, we show that, under a set of regularity conditions, the property (12) holds. We consider that $\upsilon N^{-\varrho} \leqslant n/N \leqslant \lambda$, where λ, υ and ϱ are constants such that $\lambda < 1$, $\upsilon > 0$ and $0 \leqslant \varrho < 1/2$. We assume that the π_i and $c_i^*(\psi_N)$ are such that the following regularity conditions (Berger & De La Riva Torres, 2016), hold for $\psi_N = (\theta_N^{\mathrm{T}}, \nu_N^{\mathrm{T}})^{\mathrm{T}}$:

$$\bar{\pi} \max_{i \in s}(\pi_i^{-1}) = O_p(1), \tag{13}$$

$$N^{-1} \parallel \hat{C}_\pi^*(\psi_N) - C^* \parallel = O_p(n^{-1/2}), \tag{14}$$

$$\max_{i \in s} \parallel c_i^*(\psi_N) \parallel = o_p(n^{1/2}), \tag{15}$$

$$\parallel \hat{S}^*(\psi_N) \parallel = O_p(1), \quad \parallel \hat{S}^*(\psi_N)^{-1} \parallel = O_p(1), \tag{16}$$

$$n^{-1} \bar{\pi}^{\tau} \sum_{i \in s} \left\{ \parallel c_i^*(\psi_N) \parallel \pi_i^{-1} \right\}^{\tau} = O_p(1) \quad (\tau = 2, 3, 4), \tag{17}$$

where

$$\hat{C}_\pi^*(\psi_N) = \sum_{i \in s} c_i^*(\psi_N) \pi_i^{-1}, \quad \hat{S}^*(\psi_N) = -N^{-1} \bar{\pi} \sum_{i \in s} c_i^*(\psi_N)\, c_i^*(\psi_N)^{\mathrm{T}} \pi_i^{-2}.$$

The quantities $c_i^*(\psi)$ and C^* are defined in (8). Here, $\parallel \cdot \parallel$ denotes the Frobenius norm. We consider that $c_i^*(\psi)$ is differentiable with respect to ν for all $i \in s$ in a neighbourhood around the true population value ν_N.

Condition (13) can be found in Krewski & Rao (1981, p.1014) and guarantees that π_i and $\bar{\pi}$ are of the same order of magnitude. Condition (14) assumes that $\hat{C}_\pi^*(\psi_N)$ is $\sqrt{n}$ design-consistent, and can be justified by using the Isaki & Fuller (1982, p.91) sufficient conditions. Chen & Sitter (1999) showed that (15) holds for most unequal probability sampling designs. In the independent and identically distributed setting, (15) is the consequence of the Borel–Cantelli lemma (Owen, 2001, Lemma 11.2). Condition (15) is a generalization for unequal probabilities. We need to include the constant $\bar{\pi}$ within the definition of $c_i^*(\psi_N)$ to ensure that (16) holds, but $\bar{\pi}$ can

be omitted for the computation of the function (11). Condition (17) ensures the existence of moments (e.g., Krewski & Rao, 1981).

It can be shown that the following simpler set of sufficient and stronger conditions,

$$\bar{\pi}^{-1} \max_{i \in U} \pi_i = O(1), \tag{18}$$

$$n^{-1} \sum_{i \in s} \| g_i(\psi_N) \|^{\tau} = O_p(1), \qquad n^{-1} \sum_{i \in s} \| f_i(\varphi_N) \|^{\tau} = O_p(1), \tag{19}$$

$$n^{(1-\zeta)\phi_1} N^{-1} \sum_{i \in U} \| g_i(\psi_N) \|^{2\zeta} = O(1), \qquad n^{(1-\zeta)\phi_2} N^{-1} \sum_{i \in U} \| f_i(\varphi_N) \|^{2\zeta} = O(1) \tag{20}$$

imply (13), (14), (16) and (17), where $\tau = 1, 2, 3, 4$, $\zeta = 1, 2$ and $0 \leqslant \phi_1, \phi_2 < 1$. The proof may be found in an unpublished 2016 technical report available from the second author. The conditions (18) ensure that π_i are in the vicinity of $\bar{\pi}$. The conditions (19) and (20) are conditions on the sample and population moments, which hold when $\| g_i(\psi_N) \|$ and $\| f_i(\varphi_N) \|$ are both bounded by constants or when the distributions of the $\| g_i(\psi_N) \|$ and $\| f_i(\varphi_N) \|$ do not have heavy tails.

5·2. *Asymptotic properties of the empirical likelihood point estimator*

THEOREM 1. *Under* (13)–(17)*, and for any* ψ *such that*

$$n^{-1} \bar{\pi}^{2} \sum_{i \in s} \left\{ \| g_i(\psi) \| \pi_i^{-1} \right\}^2 = O_p(1), \tag{21}$$

we have that $\hat{G}(\psi, \varphi_N)$ *in* (10) *is approximately equal to the regression estimator* (23)*; that is,*

$$N^{-1} \hat{G}(\psi, \varphi_N) = N^{-1} \hat{G}_{\mathrm{reg}}(\psi, \varphi_N) + o_p(n^{-1/2}), \tag{22}$$

with

$$\hat{G}_{\mathrm{reg}}(\psi, \varphi_N) = \hat{G}_{\pi}(\psi) - \hat{B}(\psi, \varphi_N)^{\mathrm{T}} \, \hat{f}_{\pi}(\varphi_N); \tag{23}$$

where $\hat{B}(\psi, \varphi_N)$ *is a matrix of regression coefficients defined by*

$$\hat{B}(\psi, \varphi_N) = \hat{\mathrm{var}}\{\hat{f}_{\pi}(\varphi_N)\}^{-1} \hat{\mathrm{cov}}\{\hat{f}_{\pi}(\varphi_N), \hat{G}_{\pi}(\psi)\}.$$

Here,

$$\hat{f}_{\pi}(\varphi_N) = \sum_{i \in s} f_i(\varphi_N) \pi_i^{-1}, \quad \hat{G}_{\pi}(\psi) = \sum_{i \in s} g_i(\psi) \pi_i^{-1}$$

$$\hat{\mathrm{var}}\{\hat{f}_{\pi}(\varphi_N)\} = \sum_{i \in s} \breve{f}_i(\varphi_N) \breve{f}_i(\varphi_N)^{\mathrm{T}} - \sum_{i \in s} \breve{f}_i(\varphi_N) \breve{z}_i^{\mathrm{T}} \Big(\sum_{i \in s} \breve{z}_i \breve{z}_i^{\mathrm{T}} \Big)^{-1} \sum_{i \in s} \breve{z}_i \breve{f}_i(\varphi_N)^{\mathrm{T}}, \tag{24}$$

$$\hat{\mathrm{cov}}\{\hat{f}_{\pi}(\varphi_N), \hat{G}_{\pi}(\psi)\} = \sum_{i \in s} \breve{f}_i(\varphi_N) \breve{g}_i(\psi)^{\mathrm{T}} - \sum_{i \in s} \breve{f}_i(\varphi_N) \breve{z}_i^{\mathrm{T}} \Big(\sum_{i \in s} \breve{z}_i \breve{z}_i^{\mathrm{T}} \Big)^{-1} \sum_{i \in s} \breve{z}_i \breve{g}_i(\psi)^{\mathrm{T}},$$

$$\breve{f}_i(\varphi_N) = f_i(\varphi_N) \pi_i^{-1}, \quad \textit{and} \quad \breve{g}_i(\psi_N) = g_i(\psi_N) \pi_i^{-1}.$$

The proof can be found in the Supplementary Material.

The matrix $\hat{\mathrm{var}}\{\hat{f}_{\pi}(\varphi_N)\}$ is the stratified Hansen & Hurwitz (1943) variance estimator of $\hat{f}_{\pi}(\varphi_N)$. The matrix $\hat{\mathrm{cov}}\{\hat{f}_{\pi}(\varphi_N), \hat{G}_{\pi}(\psi)\}$ is an estimator of the covariance between $\hat{G}_{\pi}(\psi)$ and $\hat{f}_{\pi}(\varphi_N)$. Expression (22) implies that the maximum empirical likelihood estimator of ψ_N

is asymptotically design-consistent. Expression (23) is the asymptotic design-optimal regression estimator proposed by Berger et al. (2003).

5·3. *Asymptotic distribution of the empirical likelihood ratio statistic*

For (12) to hold, $\hat{r}(\theta_N, \varphi_N)$ needs to converge to a quadratic form, which has a chi-squared distribution asymptotically. This is supported by the following Theorem.

THEOREM 2. *Under* (13)–(17)*, we have that* $\hat{r}(\theta_N, \varphi_N)$ *converges to a quadratic form; that is,*

$$\hat{r}(\theta_N, \varphi_N) = \hat{G}_{\mathrm{reg}}(\psi_N, \varphi_N)^{\mathrm{T}}\, \hat{Q}(\psi_N, \varphi_N)\, \hat{G}_{\mathrm{reg}}(\psi_N, \varphi_N) + O_p(n^{-1/2}), \tag{25}$$

where

$$\hat{Q}(\psi_N, \varphi_N) = (I_b - \hat{A}^{\circ})\, \hat{V}_{\mathrm{reg}}^{-1}.$$

Here, I_b *denotes the* $b \times b$ *identity matrix and*

$$\begin{aligned}
\hat{A}^{\circ} &= \hat{V}_{\mathrm{reg}}^{-1/2}\, \hat{\nabla}_g^{\circ}\, \big(\hat{\nabla}_g^{\circ\mathrm{T}}\, \hat{V}_{\mathrm{reg}}^{-1}\, \hat{\nabla}_g^{\circ}\big)^{-1} \hat{\nabla}_g^{\circ\mathrm{T}}\, \hat{V}_{\mathrm{reg}}^{-1/2}, \\
\hat{\nabla}_g^{\circ} &= \sum_{i \in s} \frac{\partial \breve{g}_i^{\circ}(\psi)}{\partial \nu}\Big|_{\psi=\psi_N}, \\
\breve{g}_i^{\circ}(\psi) &= \breve{g}_i(\psi) - \hat{B}(\psi, \varphi_N)^{\mathrm{T}}\, \breve{f}_i(\varphi_N).
\end{aligned}$$

The definition of $\breve{g}_i(\psi)$*,* $\breve{f}_i(\varphi_N)$ *and* $\hat{B}(\psi, \varphi_N)$ *can be found in Theorem* 1*. The random matrix* $\hat{V}_{\mathrm{reg}}$ *is the stratified Hansen & Hurwitz (1943) variance estimator of the regression estimator* $\hat{G}_{\mathrm{reg}}(\psi_N, \varphi_N)$*; that is,*

$$\hat{V}_{\mathrm{reg}} = \sum_{i \in s} \breve{g}_i^{\circ}(\psi_N)\, \breve{g}_i^{\circ}(\psi_N)^{\mathrm{T}} - \sum_{i \in s} \breve{g}_i^{\circ}(\psi_N)\, \breve{z}_i^{\mathrm{T}} \Big(\sum_{i \in s} \breve{z}_i\, \breve{z}_i^{\mathrm{T}} \Big)^{-1} \sum_{i \in s} \breve{z}_i\, \breve{g}_i^{\circ}(\psi_N)^{\mathrm{T}}. \tag{26}$$

The proof can be found in the Supplementary Material.

For the property (12) to hold, we need to show that the quadratic form in (25) follows a chi-squared distribution asymptotically. This can be achieved by assuming that the sampling design is such that

$$\hat{V}_{\mathrm{reg}}^{-1/2}\, \hat{G}_{\mathrm{reg}}(\psi_N, \varphi_N) \sim \mathcal{N}(0_b, I_b) \tag{27}$$

holds. Evidence for the normality of the regression estimator can be found in Scott & Wu (1981). Under sampling with replacement, standard large-sample theory can be used to show (27) (Prášková & Sen, 2009). By using expression (25) and condition (27), the random variable $\hat{r}(\theta_N, \varphi_N)$ given by expression (11) follows asymptotically a chi-squared distribution with p degrees of freedom, because $(I_b - \hat{A}^{\circ})$ is a symmetric idempotent matrix with trace p, where p is the dimension of θ. Hence, property (12) holds.

6. STRATIFIED CLUSTERED POPULATION

The population may be subdivided into a large number M of small disjoint finite subsets $\tilde{U}_i$ $(i = 1, \ldots, M)$ called clusters. Consider a stratified with-replacement sample s of n clusters selected with unequal probabilities. For example, clusters may be selected with probabilities proportional to their sizes. Units are sampled within the selected clusters. Let ψ_N be the unique solution to the equation (1), which can be re-written as $\sum_{i=1}^{M} g_{i\cdot}(\psi) = 0_b$, where $g_{i\cdot}(\psi) = \sum_{j \in \tilde{U}_i} g_{ij}(\psi)$. Here, $g_{ij}(\psi) = g_{ij}(v_{ij}, \psi)$ is the estimating function for a unit $j \subset \tilde{U}_i$, where v_{ij} is the corresponding vector of variables defining ψ_N.

We propose to use an ultimate cluster approach (e.g., Hansen et al., 1953) described as follows. Let $\hat{g}_{i\cdot}(\psi)$ be an unbiased estimator of $g_{i\cdot}(\psi)$ for a given ψ. The approach proposed in the previous sections can be used by treating the clusters as sampling units. That is, we substitute $g_i(\psi)$ by $\hat{g}_{i\cdot}(\psi)$ in (8). Now, p_i is the selection probability of the ith cluster. With population-level information, the $f_i(\varphi_N)$ in (4), are defined at cluster-level. When φ_N is a function of unit-level variables, the $f_i(\varphi_N)$ are replaced by unbiased estimates (e.g., Estevao & Särndal, 2006).

We assume that the regularity conditions (14)–(17) hold with $g_i(\psi_N)$ replaced by $\hat{g}_{i\cdot}(\psi_N)$. The equation (25) shows that $\hat{r}(\theta_N, \varphi_N)$ is approximated by a quadratic form with an ultimate cluster covariance matrix which is consistent as long as $n/M = o(1)$ (e.g., Särndal et al., 1992, Ch.4). Hence, $\hat{r}(\theta_N, \varphi_N)$ follows a chi-squared distribution asymptotically as $n \to \infty$.

7. SIMULATION STUDY

7·1. *Approaches considered*

We compare the Monte Carlo performance of the empirical likelihood confidence intervals with those based on Wald's parametric test statistic, the Q-weighted approaches, pseudo-likelihood and the rescaled bootstrap (Rao et al., 1992). The linearization approach (Binder, 1983) is omitted, because with the models considered, pseudo-likelihood or Q-weighted approaches reduce to linearization.

Q-weighted approaches are based on a Q-weighted estimator of the estimating equation (1). Q-weighted confidence intervals are based on linearization (Binder, 1983). We consider two versions. The Q-weighted 1 confidence interval is based on the design-based Hartley & Rao (1962) variance estimator. The Q-weighted 2 approach uses the conditional variance estimator (Pfeffermann & Sverchkov, 2003). The Wald and the Q-weighted approaches do not take into account the population-level information. Pseudo-likelihood confidence intervals rely on a variance estimator of an estimating equation, for a given value of θ. Binder & Patak (1994) mentioned two versions, denoted pseudo-likelihood 1 and 2. For the pseudo-likelihood 1 confidence interval, we replace ψ_N in the variance estimator by its estimate. With the pseudo-likelihood 2 approach, the nuisance parameter is kept as a function of θ. The rescaled bootstrap consists in selecting $B = 1000$ bootstrap samples of size $m = n - 1$. The quantiles of the 1000 bootstrap values are used for the confidence intervals.

We consider 95% confidence intervals. The standardized length is the average length of the confidence interval divided by $\{2 \times 1{\cdot}96\sqrt{\text{MSE}(\hat{\theta})}\}$, where $\text{MSE}(\hat{\theta})$ is the Monte Carlo mean square error of the point estimator. The ratio of average lengths is average length divided by that of the empirical likelihood confidence intervals. We also consider the standard deviation of the confidence intervals' lengths. The ratio of standard deviations is the standard deviation of the lengths divided by that of the lengths of the empirical likelihood confidence intervals. Shapiro & Wilk's (1965) test is used to assess the normality of the point estimators. In all cases, we selected 1000 random samples by using the randomized systematic sampling design with inclusion probabilities π_i and a single stratum. We used the Hartley & Rao (1962) variance estimator for approaches requiring variance estimates.

7·2. *Linear regression*

We generate a population of size $N = 10{,}000$ according to the model proposed by Hansen, Madow & Tepping (1983),

$$y_i \mid x_i \sim \Gamma\{0{\cdot}04x_i^{-3/2}(8 + 5x_i)^2, 1{\cdot}25x_i^{3/2}(8 + 5x_i)^{-1}\}, \quad \text{with } x_i \sim \Gamma(2, 5),$$

Table 1. *Hansen et al. (1983) population. 95% confidence intervals of the slope of* (28). *Shapiro & Wilk p-value* $\geqslant 0{\cdot}49$. $n = 500$. 1000 *randomized systematic samples*

	Observed coverages %	Lower tail error rates %	Upper tail error rates %	Standardised length	Ratio average lengths	Ratio standard deviations
Empirical likelihood	94·8	3·1	2·1	0·98	1·00	1·00
Wald	76·6*	23·8*	0·1*	0·63	0·96	0·53
Q-weighted 1	95·7	3·0	1·3*	0·99	0·86	0·63
Q-weighted 2	96·2	2·7	1·1*	1·03	0·89	0·64
Pseudo-likelihood 1	94·0	3·5*	2·5	0·95	0·97	1·07
Pseudo-likelihood 2	94·8	3·3	1·9	0·97	0·99	1·09
Bootstrap	96·5*	2·4	1·1*	1·03	1·05	0·91

$*$: Coverages or tail error rates significantly different respectively from 95% or $2{\cdot}5\%$, p-value $\leqslant 0{\cdot}05$.

Table 2. *Linear model with outlying values. 95% confidence intervals of the slope. Shapiro & Wilk p-value* $= 0{\cdot}057$. $n = 500$. 1000 *randomized systematic samples*

	Observed coverages %	Lower tail error rates %	Upper tail error rates %	Standardised length	Ratio average lengths	Ratio standard deviations
Empirical likelihood	95·0	2·5	2·5	1·05	1·00	1·00
Wald	94·1	2·8	3·1	0·96	0·91	0·33
Q-weighted 1	93·3*	4·9*	1·8	0·97	0·93	0·69
Q-weighted 2	93·7	4·6*	1·7	1·00	0·95	0·71
Pseudo-likelihood 1	93·3*	4·9*	1·8	0·97	0·93	0·69
Bootstrap	94·1	4·1*	1·8	1·00	0·95	0·71

$*$: Coverages or tail error rates significantly different respectively from 95% or $2{\cdot}5\%$, p-value $\leqslant 0{\cdot}05$.

where $\Gamma(\cdot\,,\cdot)$ denotes the gamma distribution. The sample size is $n = 500$. The π_i are proportional to $z_i = 5 + y_i + x_i + \epsilon_i$, where ϵ_i are generated from the unit exponential distribution.

The following linear regression model is fitted to the sample data,

$$y_i = \nu + \theta x_i + \sigma_i e_i, \tag{28}$$

where $\sigma_i = x_i^{3/4}$ and the e_i denote the residuals. Thus,

$$g_i(\psi) = (1, x_i)^{\mathrm{T}}(y_i - \nu - \theta x_i)\sigma_i^{-2}, \tag{29}$$

where $\psi = (\theta, \nu)^{\mathrm{T}}$. Let $\psi_N = (\theta_N, \nu_N)^{\mathrm{T}}$ be the solution to (1) with $g_i(\psi)$ given by (29).

The parameter of interest is the slope θ_N, and the intercept ν_N is the nuisance parameter. In Table 1, we do not reject the normality of the point estimators, because the Shapiro & Wilk test p-values are greater than $0{\cdot}49$. The coverages of θ_N for Wald and bootstrap confidence intervals are significantly different from 95%. The Q-weighted confidence intervals are the most stable, because they have a smaller ratio of standardised length, but the upper tail error rates are significantly different from $2{\cdot}5\%$. The rescaled bootstrap has the widest confidence intervals on average.

In the Supplementary Material, the significance of the intercept in (28) is tested by treating the slope as the nuisance parameter. For sample sizes less than 300, the empirical likelihood test has higher rejection rates than the model-based F-test, the Wald-test, the pseudo-likelihood tests and the Q-weighted tests.

7·3. *Linear regression with outlying values*

We consider a population of size $N = 10{,}000$. We generate population values by using $y_i = 1 + x_i + \sigma e_i$, where $x_i \sim N(8,1)$, $e_i \sim N(0,1)$, and $\sigma = 0{\cdot}75$. We replace 5% of the y_i by small values generated randomly from the uniform distribution $U(\min_{i\in U} y_i, u_1)$, with $u_1 = y_{(0{\cdot}25)} - 1{\cdot}5(y_{(0{\cdot}75)} - y_{(0{\cdot}25)})$, where $y_{(0{\cdot}25)}$ and $y_{(0{\cdot}75)}$ are respectively the lower and upper quartiles of y_i $(i \in U)$. We replace 5% of the y_i by large values generated from $U(u_2, \max_{i\in U} y_i)$, with $u_2 = y_{(0{\cdot}25)} + 1{\cdot}5(y_{(0{\cdot}75)} - y_{(0{\cdot}25)})$. The inclusion probabilities π_i are equal to $\bar{\pi}$.

The parameter ψ_N is the solution to (1) where $g_i(\psi)$ given by (29) with $\sigma_i = 1$. The slope is the parameter of interest. Table 2 shows that the empirical likelihood approach gives the correct coverages and tail error rates. The other confidence intervals have marginally lower coverages, but are slightly shorter and more stable. The distribution of the point estimator departs from normality, because the Shapiro & Wilk p-value is $0{\cdot}057$. This explains the lower coverages of the alternative approaches. The pseudo-likelihood 2 approach is omitted from Table 2, because the pseudo-likelihood 2 confidence intervals did not exist for some samples (Godambe & Thompson, 2009, p.92).

7·4. *U.K. Labour Force Survey*

We apply our approach to the first quarter 2011 U.K. Labour Force Survey, which contains data for 16–60 year-old females and 16–65 year-old males. We quadrupled the dataset to create an artificial population of size $N = 13{,}048$. The π_i are proportional to the reciprocal of the survey weights provided in the dataset. The variable y_i is the binary variable: $y_i = 1$ if the individual i is unemployed for one year or more; $y_i = 0$ otherwise. The variable x_i specifies the gender, $x_i = 1$ for male and $x_i = 0$ for female. We consider the logistic regression model with the response variable y_i and one explanatory variable, x_i,

$$g_i(\psi) = (1, x_i)^{\mathrm{T}} \left[y_i - \exp(\nu + \theta x_i)\{1 + \exp(\nu + \theta x_i)\}^{-1} \right] \tag{30}$$

where $\psi = (\theta, \nu)^{\mathrm{T}}$. Let $\psi_N = (\theta_N, \nu_N)^{\mathrm{T}}$ be the solution to (1) with $g_i(\psi)$ given by (30). The parameter of interest is the slope θ_N.

In Table 3, the coverages are similar and not significantly different from 95%. The rescaled bootstrap confidence intervals are less stable. The Q-weighted 1 approach is the same as linearization (Binder, 1983), because the same variance estimator is used and the point estimators are the same when x_i is a binary variable. Population-level information is considered in the Supplementary Material. Similar coverages and tail error rates are observed.

8. DISCUSSION

There are analogies between empirical likelihood and calibration (Deville & Särndal, 1992), although they differ. Empirical likelihood gives survey weights (6), which are naturally calibrated because of the maximisation of empirical log-likelihood function (9), and the fact that a known population parameter is fixed within the function (9). The empirical likelihood approach does not always require population-level information. With the calibration approach, the calibration distance function is only used to derive calibration weights for point estimation, and plays no role in testing or constructing confidence intervals. The empirical log-likelihood (9) is used for point estimation, testing and confidence intervals. Calibration relies on linearized variance estimates. Variance estimation is not needed for empirical likelihood. The empirical likelihood weights are positive and asymptotically optimal. Calibration weights can be negative and not necessarily asymptotically optimal.

Table 3. *U.K. Labour Force Survey data. 95% confidence intervals of the slope of the logistic model. Mean squared error of the point estimator* $= 0{\cdot}03$. *Shapiro & Wilk p-value* $> 0{\cdot}6$. $n = 600$. 1000 *randomized systematic samples.*

	Observed coverages %	Lower tail error rates %	Upper tail error rates %	Standardised length	Ratio average lengths	Ratio standard deviations
Empirical likelihood	94·9	2·4	2·7	1·01	1·00	1·00
Wald	94·2	3·4	2·4	1·00	0·97	0·92
Q-weighted 1	94·4	2·5	3·1	0·99	0·98	0·97
Q-weighted 2	95·1	2·2	2·7	1·01	1·00	0·99
Pseudo-likelihood 1	94·2	3·0	2·8	0·99	0·98	1·01
Pseudo-likelihood 2	94·3	2·9	2·8	0·99	0·98	1·04
Bootstrap	94·6	2·4	3·0	1·01	1·00	1·86

ACKNOWLEDGEMENT

We wish to thank the anonymous reviewers, Paul Smith and Professor Li-Chun Zhang for helpful comments. This research was supported by the Economic and Social Research Council, United Kingdom.

SUPPLEMENTARY MATERIAL

Supplementary Material available at *Biometrika* online includes the proof Theorem 1 in Appendix 2, the proof of Theorem 2 in Appendix 3, additional simulation studies in Appendix 4 and examples of estimating equations in Appendix 5.

APPENDIX 1

An algorithm for computing the empirical likelihood ratio statistic

As the $\hat{m}_i^*(\psi)$ maximize $\ell(m)$ under the constraint (7), for a given $\psi = (\theta^{\mathrm{T}}, \nu^{\mathrm{T}})^{\mathrm{T}}$, we have that $\hat{m}_i^*(\psi) = \{\pi_i + \overset{*}{\eta}(\psi)^{\mathrm{T}} c_i^*(\psi)\}^{-1}$, as in expression (6), where $\overset{*}{\eta}(\psi)$ is such that constraint (7) holds. Equivalently, $\overset{*}{\eta}(\psi)$ is the solution to

$$\Upsilon_1(\eta, \nu) = \sum_{i \in s} \left\{\pi_i + \overset{*}{\eta}(\psi)^{\mathrm{T}} c_i^*(\psi)\right\}^{-1} c_i^*(\psi) - C^* = 0_{r+b}. \tag{A.1}$$

Let $\ell(\theta, \nu, \varphi_{_N}) = \ell(\psi, \varphi_{_N})$. By using (9), we have

$$\ell(\theta, \nu, \varphi_{_N}) = -\sum_{i \in s} \log\{\pi_i + \overset{*}{\eta}(\psi)^{\mathrm{T}} c_i^*(\psi)\}. \tag{A.2}$$

In order to compute (11), we need to maximize (A.2) over ν. Let $\mathring{\nu}(\theta)$ be the vector ν that maximizes (A.2) for a given value of θ. As $c_i^*(\psi)$ is assumed to be differentiable with respect to ν, the vector $\mathring{\nu}(\theta)$ is the solution to

$$\frac{\partial \ell(\theta, \nu, \varphi_{_N})}{\partial \nu} = \frac{\partial \overset{*}{\eta}(\psi)^{\mathrm{T}}}{\partial \nu} \sum_{i \in s} \hat{m}_i^*(\psi)\, c_i^*(\psi) + \Upsilon_2\{\overset{*}{\eta}(\psi), \nu\} = 0_q, \tag{A.3}$$

with $\Upsilon_2\{\overset{*}{\eta}(\psi), \nu\} = \overset{*}{\eta}(\psi)^{\mathrm{T}} \sum_{i \in s} \hat{m}_i^*(\psi) \partial c_i^*(\theta, \nu)/\partial \nu$. Here, $c_i^*(\theta, \nu) = c_i^*(\psi)$, with $\psi = (\theta^{\mathrm{T}}, \nu^{\mathrm{T}})^{\mathrm{T}}$. Equation (A.3) reduces to

$$\Upsilon_2\{\overset{*}{\eta}(\psi), \nu\} = 0_q, \tag{A.4}$$

because $\sum_{i\in s}\hat{m}_i^*(\psi)\, c_i^*(\psi) = C^*$, as the $\hat{m}_i^*(\psi)$ satisfy constraint (A.1) and $\overset{*}{\eta}(\psi)^{\mathrm{T}}\, C^* = 0$, see Lemma 1 in Appendix 3.

Let $\overset{\circ}{\nu} = \overset{\circ}{\nu}(\theta)$ and $\overset{\circ}{\eta} = \overset{*}{\eta}(\overset{\circ}{\psi})$ with $\overset{\circ}{\psi} = \{\theta^{\mathrm{T}}, \overset{\circ}{\nu}(\theta)^{\mathrm{T}}\}^{\mathrm{T}}$. By definition, the vectors $\overset{\circ}{\eta}$ and $\overset{\circ}{\nu}$ satisfy the equations (A.1) and (A.4). In other words, $\overset{\circ}{\eta}$ and $\overset{\circ}{\nu}$ are the solutions to

$$\Upsilon(\eta, \nu) = 0_{r+b+q}, \tag{A.5}$$

where $\Upsilon(\eta, \nu) = \{\Upsilon_1(\eta, \nu)^{\mathrm{T}}, \Upsilon_2(\eta, \nu)^{\mathrm{T}}\}^{\mathrm{T}}$.

A root-search algorithm, such as the Newton–Raphson algorithm can be used to solve equation (A.5). This algorithm is based on the Taylor approximation of $\Upsilon(\eta, \nu)$ in the neighbourhood of $(\eta_t^{\mathrm{T}}, \nu_t^{\mathrm{T}})^{\mathrm{T}}$:

$$\Upsilon(\eta, \nu) - \Upsilon(\eta_t, \nu_t) \doteq \hat{\nabla}(\eta_t, \nu_t) \begin{pmatrix} \eta - \eta_t \\ \nu - \nu_t \end{pmatrix}, \tag{A.6}$$

where

$$\hat{\nabla}(\eta, \nu) = \partial\Upsilon(\eta, \nu)/\partial(\eta^{\mathrm{T}}, \nu^{\mathrm{T}})^{\mathrm{T}}. \tag{A.7}$$

The iterative Newton–Raphson algorithm consists in combining (A.5) and (A.6) to obtain the following recursive formula.

$$\hat{\nabla}(\eta_t, \nu_t) \begin{pmatrix} \eta_{t+1} - \eta_t \\ \nu_{t+1} - \nu_t \end{pmatrix} = -\Upsilon(\eta_t, \nu_t). \tag{A.8}$$

For the first iteration ($t = 0$), $\eta_0 = 0$ and $\nu_0 = \hat{\nu}$, where $\hat{\nu}$ is the maximum empirical likelihood estimate of $\nu_{\scriptscriptstyle N}$. The solution (η_{t+1}, ν_{t+1}) to the system of equations (A.8) gives a new set of vectors used for the next iteration. We repeat this process until convergence. At convergence, we have $\overset{\circ}{\eta}$ and $\overset{\circ}{\nu}$.

Finally, by using (A.2), we have

$$\max_{\nu\in\Lambda} \ell(\theta, \nu, \varphi_{\scriptscriptstyle N}) = \ell(\theta, \overset{\circ}{\nu}, \varphi_{\scriptscriptstyle N}) = -\sum_{i\in s} \log\{\pi_i + \overset{\circ}{\eta}{}^{\mathrm{T}} c_i^*(\theta, \overset{\circ}{\nu})\}. \tag{A.9}$$

We obtain the value of $\hat{r}(\theta, \varphi_{\scriptscriptstyle N})$ by substituting (A.9) into (11).

REFERENCES

BERGER, Y. G. (2011). Asymptotic consistency under large entropy sampling designs with unequal probabilities. *Pakistan J. Statist.* **27**, 407–426.

BERGER, Y. G. & DE LA RIVA TORRES, O. (2016). An empirical likelihood approach for inference under complex sampling design. *J. R. Stat. Soc. Ser. B.* **78**, 319–341.

BERGER, Y. G., TIRARI, M. E. H. & TILLÉ, Y. (2003). Towards optimal regression estimation in sample surveys. *Aust. N. Z. J. Stat.* **45**, 319–329.

BINDER, D. A. (1983). On the variance of asymptotically normal estimators from complex surveys. *Int. Stat. Rev.* **51**, 279–292.

BINDER, D. A. & PATAK, Z. (1994). Use of estimating functions for estimation from complex surveys. *J. Amer. Statist. Assoc.* **89**, 1035–1043.

CHAUDHURI, S., HANDCOCK, M. S. & RENDALL, M. S. (2008). Generalized linear models incorporating population level information: An empirical-likelihood-based approach. *J. R. Stat. Soc. Ser. B.* **70**, 311–328.

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.

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

CHEN, S. & KIM, J. K. (2014). Population empirical likelihood for nonparametric inference in survey sampling. *Statist. Sinica* **24**, 335–355.

DEVILLE, J. C. & SÄRNDAL, C.-E. (1992). Calibration estimators in survey sampling. *J. Amer. Statist. Assoc.* **87**, 376–382.

DURBIN, J. (1953). Some results in sampling theory when the units are selected with unequal probabilities. *J. R. Stat. Soc. Ser. B.* **15**, 262–269.

ESTEVAO, V. M. & SÄRNDAL, C.-E. (2006). Survey estimates by calibration on complex auxiliary information. *Int. Stat. Rev.* **74**, 127–147.

GABLER, S. (1984). On unequal probability sampling: Sufficient conditions for the superiority of sampling without replacement. *Biometrika* **71**, 171–175.

GODAMBE, V. P. & THOMPSON, M. (2009). Estimating functions and survey sampling. In *Sample Surveys: Inference and Analysis*, D. Pfeffermann & C. Rao, eds., Handbook of Statistics. Amsterdam: Elsevier, pp. 83–101.

GODAMBE, V. P. & THOMPSON, M. E. (1986). Parameters of superpopulation and survey population: their relationships and estimation. *Int. Stat. Rev.* **54**, 127–138.

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

HANSEN, M., HURWITZ, W. & MADOW, W. (1953). *Sample Survey Methods and Theory, volume I*. New York: John Wiley and Sons.

HANSEN, M. H. & HURWITZ, W. N. (1943). On the theory of sampling from finite populations. *Ann. of Math. Stat.* **14**, pp. 333–362.

HANSEN, M. H., MADOW, W. G. & TEPPING, B. J. (1983). An evaluation of model-dependent and probability-sampling inferences in sample surveys. *J. Amer. Statist. Assoc.* **78**, 776–793.

HARTLEY, H. O. & RAO, J. N. K. (1962). Sampling with unequal probabilities without replacement. *Ann. math. Statist. Assoc.* **33**, 350–374.

HARTLEY, H. O. & RAO, J. N. K. (1968). A new estimation theory for sample surveys. *Biometrika* **55**, 547–557.

HARTLEY, H. O. & RAO, J. N. K. (1969). A new estimation theory for sample surveys, II. In *New Developments in survey Sampling*, N. L. Johnson & H. J. Smith, eds. New York: John Wiley and Sons, pp. 147–169.

ISAKI, C. T. & FULLER, W. A. (1982). Survey design under the regression super-population model. *J. Amer. Statist. Assoc*. **77**, 89–96.

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

KREWSKI, D. & RAO, J. N. K. (1981). Inference from stratified sample: properties of linearization jackknife, and balanced repeated replication methods. *Ann. Statist.* **9**, 1010–1019.

OWEN, A. B. (1991). Empirical likelihood for linear models. *Ann. Statist.* **19**, 1725–1747.

OWEN, A. B. (2001). *Empirical Likelihood*. New York: Chapman & Hall.

PFEFFERMANN, D. & SVERCHKOV, M. (1999). Parametric and semi-parametric estimation of regression models fitted to survey data. *The Indian Journal of Statistics* Special Issue on Sample Surveys **61**, 166–186.

PFEFFERMANN, D. & SVERCHKOV, M. (2003). Fitting generalized linear models under informative probability sampling. In *Analysis of Survey Data*, R. L. Chambers & C. J. Skinner, eds. New York: John Wiley and Sons, pp. 175–195.

PRÁŠKOVÁ, Z. & SEN, P. K. (2009). Asymptotic in finite population sampling. In *Sample Surveys: Design, Methods and Applications*, D. Pfeffermann & C. Rao, eds., Handbook of Statistics. Amsterdam: Elsevier, pp. 489–522.

QIN, J. & LAWLESS, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22**, pp. 300–325.

RAO, J. N. K., WU, C. F. J. & YUE, K. (1992). Some recent work on resampling methods for complex surveys. *Survey Methodology* **18**, 209–217.

SÄRNDAL, C.-E., SWENSSON, B. & WRETMAN, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

SCOTT, A. & WU, C. (1981). On the asymptotic distribution of ratio and regression estimators. *J. Amer. Statist. Assoc*. **76**, 98–102.

SHAPIRO, S. S. & WILK, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika* **52**, pp. 591–611.

TAN, Z. & WU, C. (2015). Generalized pseudo empirical likelihood inferences for complex surveys. *Canad. J. Statist.* **43**, 117.

WU, C. & RAO, J. N. K. (2006). Pseudo-empirical likelihood ratio confidence intervals for complex surveys. *Canad. J. Statist.* **34**, 359–375.

ZHONG, B. & RAO, J. N. K. (2000). Empirical likelihood inference under stratified random sampling using auxiliary population information. *Biometrika* **87**, 929–938.

[*Received April* $20xx$*. Revised September* $20xx$]