

Empirical likelihood inference with public-use survey data

Puying Zhao

*Yunnan Key Laboratory of Statistical Modelling and Data Analysis
Yunnan University, Kunming 650091, China
e-mail: pyzhao@live.cn*

J. N. K. Rao

*School of Mathematics and Statistics, Carleton University
1125 Colonel Drive, Ottawa, Ontario K1S 5B6, Canada
e-mail: jrao34@rogers.com*

and

Changbao Wu

*Department of Statistics and Actuarial Science, University of Waterloo
200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada
e-mail: cbwu@uwaterloo.ca*

Abstract: Public-use survey data are an important source of information for researchers in social sciences and health studies to build statistical models and make inferences on the target finite population. This paper presents two general inferential tools through the pseudo empirical likelihood and the sample empirical likelihood methods. Theoretical results on point estimation and linear or nonlinear hypothesis tests involving parameters defined through estimating equations are established, and practical issues with the implementation of the proposed methods are discussed. Results from simulation studies and an application to the 2016 General Social Survey dataset of Statistics Canada show that the proposed methods work well under different scenarios. The inferential procedures and theoretical results presented in the paper make the empirical likelihood a practically useful tool for users of complex survey data.

MSC 2010 subject classifications: Primary 62D05; secondary 62G05, 62G10.

Keywords and phrases: Auxiliary information, bootstrap, calibration weighting, design-based inference, estimating equations, hypothesis test, replication weights, survey design, variable selection.

Received March 2019.

Contents

1	Introduction	2485
2	Empirical likelihood and estimating equations for complex surveys	2487
3	Empirical likelihood inference with public-use survey data	2489

3.1	Public-use survey data and basic assumptions	2489
3.2	The pseudo empirical likelihood approach	2491
3.3	The sample empirical likelihood approach	2493
3.4	Design-based variable selection	2495
4	Bootstrap calibrated empirical likelihood methods	2496
5	Simulation studies	2498
6	An application to the GSS 2016 dataset	2502
7	Additional remarks	2505
A	Theoretical justification of the bootstrap calibration method	2506
	Acknowledgements	2507
	References	2507

1. Introduction

Owen [17] proposed the empirical likelihood approach for making inference from independent and identically distributed random samples. He showed that the empirical likelihood ratio statistic for the population mean has a standard limiting chi-squared distribution, and used this result to obtain confidence intervals for the population mean similar to the classical parametric method. The shape and orientation of the empirical likelihood ratio confidence intervals are determined entirely by the data and the intervals are range preserving and transformation invariant, unlike the normal theory intervals. Qin and Lawless [18] demonstrated that empirical likelihood can be combined with estimating equations for statistical inferences on more general parameters. The development of empirical likelihood as a general inferential tool has been one of the major advances in statistics in the past three decades.

Empirical likelihood was in fact first introduced in the sample survey context by Hartley and Rao [13] as the scale-load likelihood, but their focus was on point estimation of a finite population mean under simple random sampling and stratified simple random sampling. Chen and Qin [6] studied empirical likelihood under simple random sampling using the formulation of Owen [17], and Zhong and Rao [32] studied empirical likelihood confidence intervals on the finite population mean under stratified simple random sampling. For general sampling designs involving unequal probability sampling with or without stratification, there have been several proposed approaches on empirical likelihood for complex surveys, including the pseudo empirical likelihood method of Chen and Sitter [7] and Wu and Rao [28], the population empirical likelihood method of Chen and Kim [9], and the empirical likelihood method of Berger and Torres [3] and Oguz-Alper and Berger [16]. However, all existing methods require the first order inclusion probabilities from the initial survey design and are developed under the setting that detailed design information is available. In addition, the use of calibration constraints for inference with existing approaches requires that auxiliary information, such as variables with known population means or totals used for calibration, is available to survey data users.

In practice, public-use survey data are released to users and such data sets often report only the variables of interest and the final survey weights $\{w_i, i \in \mathcal{S}\}$ obtained by adjusting for unit nonresponse and calibration on auxiliary variables selected by the producer of the data, where \mathcal{S} denotes the set of units included in the released data file. Furthermore, the data file provides B columns of final replication weights $\{w_i^{(b)}, i \in \mathcal{S}\}$, $b = 1, \dots, B$ designed for variance estimation. The following table shows a typical format of public-use survey data files available to the users, where (y_1, y_2) are the study variables and (x_1, x_2, x_3) are the associated covariates used in the analysis.

i	y_{i1}	y_{i2}	x_{i1}	x_{i2}	x_{i3}	w_i	$w_i^{(1)}$	\dots	$w_i^{(B)}$
1	y_{11}	y_{12}	x_{11}	x_{12}	x_{13}	w_1	$w_1^{(1)}$	\dots	$w_1^{(B)}$
2	y_{21}	y_{22}	x_{21}	x_{22}	x_{23}	w_2	$w_2^{(1)}$	\dots	$w_2^{(B)}$
3	y_{31}	y_{32}	x_{31}	x_{32}	x_{33}	w_3	$w_3^{(1)}$	\dots	$w_3^{(B)}$
\vdots	\vdots	\vdots							
n	y_{n1}	y_{n2}	x_{n1}	x_{n2}	x_{n3}	w_n	$w_n^{(1)}$	\dots	$w_n^{(B)}$

The final replication weights $\{w_i^{(b)}, i \in \mathcal{S}\}$ are one of the most crucial parts in creating public-use survey data files. Different versions of bootstrap replication weights, such as those developed by Rao and Wu [22] and Rao, Wu and Yue [23] for stratified multi-stage designs, are commonly reported with the data file. Final replication weights are typically obtained by subjecting the basic replication weights (such as the bootstrap weights) to the same unit nonresponse adjustment and calibration procedures. None of the existing empirical likelihood methods is applicable for statistical inferences with public-use data files because the first order inclusion probabilities, the calibration variables and the associated known population means or totals are not reported on the data file and are not available to users. Under multistage sampling, the primary sampling units or clusters are also not identified on the data file.

The main purpose of this article is to present empirical likelihood methods for statistical analysis with public-use survey data files. We consider two general approaches: the first is based on the pseudo empirical likelihood and the second uses the sample empirical likelihood, to be described in Sections 2 and 3. We present design-based inferential procedures and theoretical results on two general statistical inference problems with the vector of finite population parameters defined through the census estimating equations: the maximum empirical likelihood estimators and the empirical likelihood ratio test on a general linear or non-linear hypothesis. Design-based variable selection through a penalized pseudo or sample empirical likelihood is discussed. We also present a bootstrap procedure under single stage survey designs for creating valid replication weights with theoretical justifications. Simulation results and an application to the General Social Survey 2016 public-use data file released by Statistics Canada are included.

The contributions of the paper include (i) discussions on the general format and framework for creating public-use data files; (ii) presentations of empirical likelihood inferential procedures for public-use datasets; (iii) theoretical justi-

fications of a bootstrap procedure under calibration weighting, which partially addresses the question of how to create replication weights for public-use data files; and (iv) demonstrations of empirical likelihood methods through simulation studies and real data analysis. Our presentation on sample empirical likelihood follows Zhao, Haziza and Wu [30], and the discussion on pseudo empirical likelihood follows Zhao and Wu [29]. Proofs and technical details of main theoretical results have similarities to Qin and Lawless [18] and Zhao, Haziza and Wu [30], and are presented in the online version of the paper (Zhao, Rao and Wu, [31]). The topics of how to deal with item nonresponse and how to create replication weights under imputation for missing data are not addressed in the paper. They belong to a very broad research area that has been actively pursued by many researchers.

The basic settings are described in Section 2. Main theoretical results on empirical likelihood methods are presented in Section 3. A bootstrap calibration procedure under single stage survey designs to create valid replication weights is described in Section 4 with theoretical justification given in the Appendix A. Results from simulation studies are reported in Section 5. An application to the General Social Survey 2016 public-use data file is presented in Section 6. We conclude with some additional remarks in Section 7.

2. Empirical likelihood and estimating equations for complex surveys

Let $\mathcal{U} = \{1, 2, \dots, N\}$ be the set of units in the finite population, where N is the population size. Let (y_i, \mathbf{x}_i) be the measures of the study variable y and auxiliary variables \mathbf{x} for unit i . Let $\mathcal{F}_N = \{(y_i, \mathbf{x}_i), i = 1, \dots, N\}$ represent the survey population and let $\{(y_i, \mathbf{x}_i), i \in \mathcal{S}\}$ be the survey sample data. Let $\pi_i = P(i \in \mathcal{S})$, $i = 1, \dots, N$ be the first order inclusion probabilities.

Survey data are a major source of information for official statistics, where the focus is often on descriptive population quantities such as population means or quantiles. Complex surveys are also frequently used by researchers in social sciences and medical and health studies for statistical modelling. Under both scenarios, the finite population parameters $\boldsymbol{\theta}_N$ of dimension p can be defined as the solution to the census estimating equations

$$U_N(\boldsymbol{\theta}) = \sum_{i=1}^N \mathbf{g}(\mathbf{x}_i, y_i, \boldsymbol{\theta}) = \mathbf{0}, \quad (2.1)$$

where $\mathbf{g}(\mathbf{x}, y, \boldsymbol{\theta})$ is a vector of estimating functions of dimension r , and $\boldsymbol{\theta} \in \Theta$, a compact subset of \mathcal{R}^p with $1 \leq p \leq r$. Under normal circumstances we have $r = p$ but over-identified scenarios with $r > p$ do arise in practice due to additional calibration constraints or known moment conditions over certain variables.

Standard empirical likelihood inference with independent observations, as introduced by Owen [17] and with parameters defined by estimating equations,

as discussed by Qin and Lawless [18], consists of three ingredients:

$$\ell(\mathbf{p}) = \sum_{i \in \mathcal{S}} \log(p_i), \quad (2.2)$$

$$\sum_{i \in \mathcal{S}} p_i = 1, \quad (2.3)$$

$$\sum_{i \in \mathcal{S}} p_i \mathbf{g}(\mathbf{x}_i, y_i, \boldsymbol{\theta}) = \mathbf{0}, \quad (2.4)$$

where $\ell(\mathbf{p})$ given by (2.2) is the empirical log-likelihood function and $\mathbf{p} = (p_1, \dots, p_n)'$ is the probability measure over the n sampled units, equation (2.3) is the normalization constraint to ensure that \mathbf{p} is a discrete probability measure, and equations (2.4) are the constraints induced by the parameters $\boldsymbol{\theta}$. The use of $\log(p_i)$ implicitly requires that $p_i > 0$.

Naive applications of the standard empirical likelihood methods to complex survey data do not produce valid results under the design-based framework. There have been three major modified approaches in the survey sampling literature on using the empirical likelihood method for complex survey data, and their relations to the standard empirical likelihood ingredients (2.2), (2.3) and (2.4) can be described as follows.

(1) *The pseudo empirical likelihood approach (PEL)*: Chen and Sitter [7] suggested to replace $\ell(\mathbf{p})$ by $\ell_{\text{PEL0}}(\mathbf{p}) = \sum_{i \in \mathcal{S}} d_i \log(p_i)$, where $d_i = \pi_i^{-1}$ are the basic design weights, while constraints (2.3) and (2.4) remain unchanged. The use of $\ell_{\text{PEL0}}(\mathbf{p})$ is motivated by the fact that $\ell_{\text{PEL0}}(\mathbf{p})$ is the Horvitz-Thompson estimator for the conceptual census empirical log-likelihood function $\sum_{i=1}^N \log(p_i)$. Wu and Rao [28] used a modified version $\ell_{\text{PEL1}}(\mathbf{p}) = n \sum_{i \in \mathcal{S}} \tilde{d}_i(\mathcal{S}) \log(p_i)$, where $\tilde{d}_i(\mathcal{S}) = d_i / \sum_{j \in \mathcal{S}} d_j$, which facilitates the construction of the pseudo empirical likelihood ratio confidence intervals for population parameters.

(2) *The population empirical likelihood approach (POEL)*: Chen and Kim [9] defined the population empirical log-likelihood function as $\ell_{\text{POEL}} = \sum_{i=1}^N \log(\omega_i)$ with normalization constraint $\sum_{i=1}^N \omega_i = 1$. The survey data and parameters are forced into the “population system” through the constraints $\sum_{i \in \mathcal{S}} \omega_i \pi_i^{-1} = 1$ and $\sum_{i \in \mathcal{S}} \omega_i \{\mathbf{g}(\mathbf{x}_i, y_i, \boldsymbol{\theta}) \pi_i^{-1}\} = \mathbf{0}$. Chen and Kim [9] focused on Poisson sampling and rejective sampling, and the method has not been developed for general unequal probability sampling designs or general inferential problems for analytical uses of survey data.

(3) *The sample empirical likelihood approach (SEL)*: The method was first mentioned very briefly by Chen and Kim [9] as a remark but detailed exploration was not pursued in their paper. The idea is to use the standard empirical log-likelihood function $\ell_{\text{SEL0}}(\mathbf{p}) = \sum_{i \in \mathcal{S}} \log(p_i)$ from (2.2) and the standard normalization constraint (2.3) but modify the constraints induced by the parameters as $\sum_{i \in \mathcal{S}} p_i \{\mathbf{g}(\mathbf{x}_i, y_i, \boldsymbol{\theta}) \pi_i^{-1}\} = \mathbf{0}$. A related formulation was presented by Berger and De La Riva Torres ([1], [2] and [3]) and Oguz-Alper and Berger [16]. They used $l_{(m)} = \sum_{i \in \mathcal{S}} \log(m_i)$, where the m_i satisfy the so-called design

constraint $\sum_{i \in \mathcal{S}} m_i \pi_i = n$. The constraints for the parameters are specified as $\sum_{i \in \mathcal{S}} m_i \mathbf{g}(\mathbf{x}_i, y_i, \boldsymbol{\theta}) = \mathbf{0}$. It can be seen that, if we let $p_i = m_i \pi_i n^{-1}$, the formulation is equivalent to the one proposed by Chen and Kim [9]. The sample empirical likelihood method has been further developed in the paper by Zhao, Haziza and Wu [30] as a general inference tool for survey data analysis under the assumption that the first order inclusion probabilities π_i and other related design and population information are available.

Unfortunately, none of the existing empirical likelihood methods can be used directly for statistical analysis with public-use survey data files since the initial inclusion probabilities π_i are not available, and calibration variables along with their known population totals are typically not given to the end users of the data files. On the other hand, the availability of final survey weights and associated replication weights for public-use datasets provides a unique opportunity to develop empirical likelihood as a general statistical tool for survey data analysis.

3. Empirical likelihood inference with public-use survey data

3.1. Public-use survey data and basic assumptions

Consider the following version of a micro survey data file, which is released by the survey agency for public use:

$$\left\{ \left(y_i, \mathbf{x}_i, w_i, w_i^{(1)}, \dots, w_i^{(B)} \right), i = 1, 2, \dots, n \right\},$$

where the y_i and \mathbf{x}_i are possibly vector-values survey variables included in the dataset, the w_i is the final survey weight for unit i after unit nonresponse adjustment and/or calibration weighting, and n is the final sample size. Also included in the data file are B final replication weights $w_i^{(1)}, \dots, w_i^{(B)}$ associated with unit i . The detailed survey design information such as the original design weights $d_i = 1/\pi_i$ and the known auxiliary population information are assumed to be unavailable to the users of the data file. It is also assumed that the finite population size N is unknown.

The survey weighted estimating equations for the vector of parameters $\boldsymbol{\theta}_N$ defined by the census estimating equations (2.1) are given by

$$\hat{U}_n(\boldsymbol{\theta}) = \sum_{i \in \mathcal{S}} w_i \mathbf{g}(\mathbf{x}_i, y_i, \boldsymbol{\theta}) = \mathbf{0}. \quad (3.1)$$

For standard scenarios where $r = p$, i.e., the number of equations is the same as the number of parameters, the survey weighted estimator $\hat{\boldsymbol{\theta}}_N$ for $\boldsymbol{\theta}_N$ is the solution to (3.1). Let $\mathbf{g}_i(\boldsymbol{\theta}) = \mathbf{g}(\mathbf{x}_i, y_i, \boldsymbol{\theta})$ and assume that $\mathbf{g}_i(\boldsymbol{\theta})$ is a smooth function of $\boldsymbol{\theta}$. The approximate design-based variance of $\hat{\boldsymbol{\theta}}_N$ has the well-known sandwich form (Binder, [4])

$$\text{Var}(\hat{\boldsymbol{\theta}}_N) \doteq \Gamma^{-1} \text{Var}\{N^{-1} \hat{U}_n(\boldsymbol{\theta}_N)\} (\Gamma^{-1})',$$

where $\Gamma = \Gamma(\boldsymbol{\theta}_N)$, $\Gamma(\boldsymbol{\theta}) = N^{-1} \sum_{i=1}^N \partial \mathbf{g}_i(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}$ and $\text{Var}\{N^{-1} \hat{U}_n(\boldsymbol{\theta}_N)\}$ is the design-based variance. There have been attempts to address hypothesis testing problems involving a single component of the vector of parameters $\boldsymbol{\theta}_N$ under the estimating equations framework, see, for instance, Binder and Patak [5], but general hypothesis testing procedures are not available in the literature.

We consider smooth estimating functions and allow over-identified estimating equations system with $r \geq p$. Practically useful results for the special case $r = p$ and for a scalar parameter (i.e., $p = 1$) will also be spelled out. For asymptotic development, we assume that there is a sequence of finite populations and a sequence of survey designs with both the population size N and the sample size n going to infinity; see Isaki and Fuller [14] for further detail. We use $N \rightarrow \infty$ to denote the limiting process. Note that $\boldsymbol{\theta}_N$ refers to the true vector of the finite population parameters. Throughout the paper, we use $\|\cdot\|$ to denote the Euclidean norm and $\xrightarrow{\mathcal{L}}$ to denote convergence in distribution under the design-based framework. Let $O_p(\cdot)$ and $o_p(\cdot)$ be the stochastic orders under the same framework. We consider the following basic assumptions for the public-use survey data file and the estimating functions $\mathbf{g}_i(\boldsymbol{\theta})$.

Assumption 1. *The final survey weights (w_1, w_2, \dots, w_n) and the finite population values $\mathcal{F}_N = \{(y_i, \mathbf{x}_i), i = 1, \dots, N\}$ satisfy conditions that ensure $\hat{U}_n(\boldsymbol{\theta}_N) = \sum_{i \in \mathcal{S}} w_i \mathbf{g}_i(\boldsymbol{\theta}_N)$ is asymptotically normally distributed with mean zero and variance-covariance matrix of the order $O(N^2/n)$.*

Let $\hat{\boldsymbol{\eta}}^{(b)}(\boldsymbol{\theta}_N) = \sum_{i \in \mathcal{S}} w_i^{(b)} \mathbf{g}_i(\boldsymbol{\theta}_N)$ be the replicate version of the initial $\hat{U}_n(\boldsymbol{\theta}_N) = \sum_{i \in \mathcal{S}} w_i \mathbf{g}_i(\boldsymbol{\theta}_N)$ using the b th set of replication weights $(w_1^{(b)}, w_2^{(b)}, \dots, w_n^{(b)})$, $b = 1, 2, \dots, B$ and treating $\boldsymbol{\theta}_N$ as a known number.

Assumption 2. *The final replication weights ensure that the replication variance estimator*

$$v\{\hat{U}_n(\boldsymbol{\theta}_N)\} = \frac{1}{B} \sum_{b=1}^B \left\{ \hat{\boldsymbol{\eta}}^{(b)}(\boldsymbol{\theta}_N) - \hat{U}_n(\boldsymbol{\theta}_N) \right\} \left\{ \hat{\boldsymbol{\eta}}^{(b)}(\boldsymbol{\theta}_N) - \hat{U}_n(\boldsymbol{\theta}_N) \right\}' \quad (3.2)$$

is a design-consistent estimator of the true but unknown variance-covariance matrix $\text{Var}\{\hat{U}_n(\boldsymbol{\theta}_N) \mid \mathcal{F}_N\}$.

The original design weights, the nonresponse adjusted weights and the calibration weights usually satisfy Assumption 1. It is part of the foundation for design-based inference. Assumption 2 is the guiding principle for public-use data file producers on how to create replication weights and for research activities on replication methods for variance estimation in surveys. Note that Assumption 2 does not necessarily require a large B for the given dataset, as shown by the results presented in Kim and Wu [15]. Most survey organizations, including Statistics Canada, use $B = 500$ for producing public-use survey data files in their current practice. See the example of General Social Survey presented in Section 6.

Assumption 3. *(i) $\lim_{N \rightarrow \infty} (n/N) = \gamma \in (0, 1)$; (ii) $c_1 < w_i N/n < c_2$, $i \in \mathcal{S}$ for some positive constants c_1 and c_2 ; (iii) $N^{-1} \sum_{i \in \mathcal{S}} w_i - 1 = O_p(n^{-1/2})$.*

Assumption 4. (i) $\sup_{\theta \in \Theta} N^{-1} \sum_{i \in \mathcal{S}} \|\mathbf{g}_i(\theta)\|^\kappa < c$ for some $\kappa > 2$ and some positive constant c ; (ii) $\max_{i \in \mathcal{S}} \sup_{\theta \in \Theta} \|\mathbf{g}_i(\theta)\| = o_p(n^{1/2})$.

Assumption 5. (i) The matrices $W_1(\theta_N) = N^{-1} \sum_{i=1}^N \mathbf{g}_i(\theta_N) \mathbf{g}_i(\theta_N)'$, $W_2(\theta_N) = nN^{-2} E\{\sum_{i \in \mathcal{S}} w_i^2 \mathbf{g}_i(\theta_N) \mathbf{g}_i(\theta_N)' \mid \mathcal{F}_N\}$ and $\Omega(\theta_N) = nN^{-2} Var\{\sum_{i \in \mathcal{S}} w_i \mathbf{g}_i(\theta_N) \mid \mathcal{F}_N\}$ are all positive definite; (ii) $\Gamma(\theta_N) = N^{-1} \sum_{i=1}^N \partial \mathbf{g}_i(\theta) / \partial \theta|_{\theta=\theta_N}$ has full column rank p .

Assumptions 3–5 are standard regularity conditions for asymptotic development for finite populations with complex survey data. The inclusion of the factors N^{-1} or N^{-2} in the quantities presented in Assumption 5 is for convenience in asymptotic orders. They are not required for computational purposes as they all cancel out in the main results to be presented in the next two subsections. It should be noted that the pseudo empirical likelihood approach of Section 3.2 and the sample empirical likelihood approach of Section 3.3 are formulated using the final weights w_i which contain partial information on the survey design and the creation of the dataset. The empirical likelihood ratio statistics for both approaches do not have standard χ^2 asymptotic distributions, and information from additional columns of replication weights in the dataset is required for computing the variance components involved.

3.2. The pseudo empirical likelihood approach

Let $\tilde{w}_i(\mathcal{S}) = w_i / \sum_{k \in \mathcal{S}} w_k$, $i \in \mathcal{S}$ be the normalized final survey weights. The pseudo empirical log-likelihood function is defined as

$$\ell_{\text{PEL}}(\mathbf{p}) = n \sum_{i \in \mathcal{S}} \tilde{w}_i(\mathcal{S}) \log(p_i).$$

Maximizing $\ell_{\text{PEL}}(\mathbf{p})$ subject to the normalization constraint (2.3), i.e., $\sum_{i \in \mathcal{S}} p_i = 1$, gives $\hat{\mathbf{p}} = (\hat{p}_1, \dots, \hat{p}_n)'$, where $\hat{p}_i = \tilde{w}_i(\mathcal{S})$. Let $\hat{\mathbf{p}}(\theta) = (\hat{p}_1(\theta), \dots, \hat{p}_n(\theta))'$ be the maximizer of $\ell_{\text{PEL}}(\mathbf{p})$ under the normalization constraint (2.3) and the parameter constraint (2.4), i.e., $\sum_{i \in \mathcal{S}} p_i \mathbf{g}_i(\theta) = \mathbf{0}$, for a fixed value of θ . It can be shown that $\hat{p}_i(\theta) = \tilde{w}_i(\mathcal{S}) / \{1 + \boldsymbol{\lambda}' \mathbf{g}_i(\theta)\}$ for $i \in \mathcal{S}$, where the Lagrange multiplier $\boldsymbol{\lambda} = \boldsymbol{\lambda}(\theta)$ is the solution to

$$\mathbf{g}_{\text{PEL}}(\boldsymbol{\lambda}) = \sum_{i \in \mathcal{S}} \frac{\tilde{w}_i(\mathcal{S}) \mathbf{g}_i(\theta)}{1 + \boldsymbol{\lambda}' \mathbf{g}_i(\theta)} = \mathbf{0}, \tag{3.3}$$

which can be solved using the modified Newton-Raphson method presented in Chen, Sitter and Wu [8] and the R code described in Wu [26]. The maximum pseudo empirical likelihood estimator $\hat{\boldsymbol{\theta}}_{\text{PEL}}$ is the maximizer of $\ell_{\text{PEL}}\{\hat{\mathbf{p}}(\theta)\} = n \sum_{i \in \mathcal{S}} \tilde{w}_i(\mathcal{S}) \log\{\hat{p}_i(\theta)\}$ with respect to θ . For the special case $r = p$, the estimator $\hat{\boldsymbol{\theta}}_{\text{PEL}}$ is the solution to

$$\sum_{i \in \mathcal{S}} \hat{p}_i \mathbf{g}_i(\theta) = \sum_{i \in \mathcal{S}} \tilde{w}_i(\mathcal{S}) \mathbf{g}_i(\theta) = \mathbf{0},$$

which is the same as the customary survey weighted estimating equations estimator $\hat{\theta}_N$.

The pseudo empirical log-likelihood ratio statistic for θ is given by

$$r_{\text{PEL}}(\theta) = \ell_{\text{PEL}}\{\hat{\boldsymbol{p}}(\theta)\} - \ell_{\text{PEL}}(\hat{\boldsymbol{p}}) = -n \sum_{i \in \mathcal{S}} \tilde{w}_i(\mathcal{S}) \log\{1 + \boldsymbol{\lambda}' \mathbf{g}_i(\theta)\}.$$

We can re-write the maximum pseudo empirical likelihood estimator of θ_N as $\hat{\theta}_{\text{PEL}} = \arg \max_{\theta \in \Theta} r_{\text{PEL}}(\theta)$. Theorem 1 presents asymptotic properties of the estimator $\hat{\theta}_{\text{PEL}}$. Note that the quantities $W_1(\theta_N)$, $\Gamma(\theta_N)$ and $\Omega(\theta_N)$ in Theorem 1 are defined in Assumption 5.

Theorem 1. *Under Assumptions 1, 3, 4 and 5, we have*

$$n^{1/2}(\hat{\theta}_{\text{PEL}} - \theta_N) \mid \mathcal{F}_N \xrightarrow{\mathcal{L}} N(0, V_1),$$

where $V_1 = \Sigma_1 \Gamma' W_1^{-1} \Omega W_1^{-1} \Gamma \Sigma_1$, $\Sigma_1 = (\Gamma' W_1^{-1} \Gamma)^{-1}$, $W_1 = W_1(\theta_N)$, $\Gamma = \Gamma(\theta_N)$ and $\Omega = \Omega(\theta_N)$.

Proofs of Theorem 1 and Theorems 2–6 presented below resemble the proofs in Zhao et al. [30]. Details are available in the online version of the paper at arXiv (Zhao et al., [31]). Proofs of Theorems 1 and 4 are also similar to the proof of Theorem 1 in Qin and Lawless [18].

Corollary 1. *Under the assumptions in Theorem 1 and $r = p$ (i.e., the number of equations is the same as the number of parameters), the asymptotic variance-covariance matrix V_1 of $\hat{\theta}_{\text{PEL}}$ reduces to $V_1 = \Gamma^{-1} \Omega (\Gamma')^{-1}$.*

Suppose we want to test the simple hypothesis: $H_0 : \theta_N = \theta_{N_0}$ against $H_1 : \theta_N \neq \theta_{N_0}$. The pseudo empirical log-likelihood ratio statistic for testing H_0 is given by

$$\text{LR}_{\text{PEL}}(\theta_{N_0}) = 2\{r_{\text{PEL}}(\hat{\theta}_{\text{PEL}}) - r_{\text{PEL}}(\theta_{N_0})\} = 2\{\ell_{\text{PEL}}(\hat{\theta}_{\text{PEL}}) - \ell_{\text{PEL}}(\theta_{N_0})\}.$$

The asymptotic distribution of $\text{LR}_{\text{PEL}}(\theta_{N_0})$ is given in Theorem 2.

Theorem 2. *Suppose that Assumptions 1, 3, 4 and 5 hold. Then*

$$\text{LR}_{\text{PEL}}(\theta_{N_0}) \mid \mathcal{F}_N \xrightarrow{\mathcal{L}} Q' \Delta_1 Q,$$

where $Q \sim N(0, I_r)$, I_r is the $r \times r$ identity matrix, r is the dimension of the estimating functions $\mathbf{g}_i(\theta)$, and $\Delta_1 = \Omega^{1/2} W_1^{-1} \Gamma \Sigma_1 \Gamma' W_1^{-1} \Omega^{1/2}$ with $\Sigma_1 = (\Gamma' W_1^{-1} \Gamma)^{-1}$.

Corollary 2. *Under the assumptions in Theorem 2 and $r = p$, we have $\Delta_1 = \Omega^{1/2} W_1^{-1} \Omega^{1/2}$. In particular, if $r = p = 1$, then*

$$\text{LR}_{\text{PEL}}(\theta_{N_0}) \mid \mathcal{F}_N \xrightarrow{\mathcal{L}} (\Omega/W_1) \chi^2(1),$$

where $\chi^2(1)$ denotes the standard χ^2 random variable with one degree of freedom.

We further consider a pseudo empirical log-likelihood ratio test for a general linear or nonlinear hypothesis $H_0: R(\boldsymbol{\theta}_N) = \mathbf{0}$ against $H_1: R(\boldsymbol{\theta}_N) \neq \mathbf{0}$, where $R(\boldsymbol{\theta}_N)$ is a $k \times 1$ vector-valued functions with $k \leq p$ and $R(\boldsymbol{\theta}_N) = \mathbf{0}$ imposes k constraints on the vector of parameters $\boldsymbol{\theta}_N$. Let $\Theta^R = \{\boldsymbol{\theta} \mid \boldsymbol{\theta} \in \Theta \text{ and } R(\boldsymbol{\theta}) = \mathbf{0}\}$ be the restricted parameter space under H_0 . The restricted maximum pseudo empirical likelihood estimator of $\boldsymbol{\theta}$ under H_0 is defined as $\hat{\boldsymbol{\theta}}_{\text{PEL}}^R = \arg \max_{\boldsymbol{\theta} \in \Theta^R} r_{\text{PEL}}(\boldsymbol{\theta})$. The pseudo empirical log-likelihood ratio statistic for testing H_0 versus H_1 is given by

$$\text{LR}_{\text{PEL}}(\boldsymbol{\theta}_N \mid H_0) = 2\{r_{\text{PEL}}(\hat{\boldsymbol{\theta}}_{\text{PEL}}) - r_{\text{PEL}}(\hat{\boldsymbol{\theta}}_{\text{PEL}}^R)\} = 2\{\ell_{\text{PEL}}(\hat{\boldsymbol{\theta}}_{\text{PEL}}) - \ell_{\text{PEL}}(\hat{\boldsymbol{\theta}}_{\text{PEL}}^R)\}.$$

Theorem 3. *Suppose that Assumptions 1, 3, 4 and 5 hold. If the function $R(\boldsymbol{\theta})$ is twice continuously differentiable and $\Phi(\boldsymbol{\theta}_N) = \partial R(\boldsymbol{\theta})/\partial \boldsymbol{\theta}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_N}$ has rank k , then*

$$\text{LR}_{\text{PEL}}(\boldsymbol{\theta}_N \mid H_0) \mid \mathcal{F}_N \xrightarrow{\mathcal{L}} Q' \Delta_1^R Q,$$

where $Q \sim N(0, I_r)$, $\Delta_1^R = \Omega^{1/2} W_1^{-1} \Gamma \Sigma_1 \Phi' (\Phi \Sigma_1 \Phi)^{-1} \Phi \Sigma_1 \Gamma' W_1^{-1} \Omega^{1/2}$ and $\Phi = \Phi(\boldsymbol{\theta}_N)$.

Let $\delta_j, j = 1, \dots, p$ be the non-zero eigenvalues of the $r \times r$ matrix Δ_1 . The asymptotic distribution of $\text{LR}_{\text{PEL}}(\boldsymbol{\theta}_N)$ given in Theorem 2 can be alternatively represented by $\sum_{j=1}^p \delta_j \chi_j^2(1)$, where $\chi_j^2(1), j = 1, \dots, p$ are independent random variables, all following the same distribution as $\chi^2(1)$. Similarly, the distribution of the quadratic form $Q^T \Delta_1^R Q$ given in Theorem 3 can be alternatively represented by $\sum_{j=1}^k \delta_j^R \chi_j^2(1)$, where $\delta_j^R, j = 1, \dots, k$ are the non-zero eigenvalues of the matrix Δ_1^R .

Practical implementations of the theoretical results generally require the estimation of the asymptotic variance V_1 for Theorem 1, the matrix Δ_1 for Theorem 2 and Δ_1^R for Theorem 3. This amounts to estimating the involved components W_1, Γ, Ω and Φ . By the simple ‘‘plug-in’’ method, we can estimate the term W_1 by $\hat{W}_1 = N^{-1} \sum_{i \in S} w_i \mathbf{g}_i(\hat{\boldsymbol{\theta}}_{\text{PEL}}) \mathbf{g}_i(\hat{\boldsymbol{\theta}}_{\text{PEL}})'$, the term Γ by $\hat{\Gamma}_{\text{PEL}} = N^{-1} \sum_{i \in S} w_i \partial \mathbf{g}_i(\boldsymbol{\theta})/\partial \boldsymbol{\theta}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\text{PEL}}}$, and estimate $\Phi = \Phi(\boldsymbol{\theta}_N)$ by $\hat{\Phi} = \Phi(\hat{\boldsymbol{\theta}}_{\text{PEL}})$. The most critical component Ω can be estimated by $\hat{\Omega}_{\text{PEL}} = nN^{-2} v\{\hat{U}_n(\hat{\boldsymbol{\theta}}_{\text{PEL}})\}$, where $v\{\hat{U}_n(\hat{\boldsymbol{\theta}}_{\text{PEL}})\}$ is the replication variance estimator outlined in Assumption 2 using the replication weights from the survey data file.

The distribution of the quadratic forms $Q' \Delta_1 Q$ and $Q' \Delta_1^R Q$ may also be approximated by the Rao-Scott (RS) correction method (Rao and Scott, [19], [20]). For instance, the first-order RS correction leads to $\text{LR}_{\text{PEL}}(\boldsymbol{\theta}_N) \mid \mathcal{F}_N \xrightarrow{\mathcal{L}} a \chi^2(p)$, where $a = \sum_{j=1}^p \delta_j/p$. The second-order RS correction gives $\text{LR}_{\text{PEL}}(\boldsymbol{\theta}_N) \mid \mathcal{F}_N \xrightarrow{\mathcal{L}} c \chi^2(k^*)$, where $c = \sum_{j=1}^p \delta_j^2 / \sum_{j=1}^p \delta_j$ and $k^* = (\sum_{j=1}^p \delta_j)^2 / \sum_{j=1}^p \delta_j^2$.

3.3. The sample empirical likelihood approach

The sample empirical likelihood approach described in Section 2 can be adapted for public-use survey data. We start with the standard empirical log-likelihood

function $\ell_{\text{SEL}}(\mathbf{p}) = \sum_{i \in \mathcal{S}} \log(p_i)$. Maximizing $\ell_{\text{SEL}}(\mathbf{p})$ under the normalization constraint (2.3), i.e., $\sum_{i \in \mathcal{S}} p_i = 1$, gives $\hat{p}_i = n^{-1}$, $i \in \mathcal{S}$. The constraints for the parameters $\boldsymbol{\theta}$ defined through (2.1) are formed using the weighted estimating functions $w_i \mathbf{g}_i(\boldsymbol{\theta})$ and are given by

$$\sum_{i \in \mathcal{S}} p_i \{w_i \mathbf{g}_i(\boldsymbol{\theta})\} = \mathbf{0}. \quad (3.4)$$

Let $\hat{\mathbf{p}}(\boldsymbol{\theta}) = (\hat{p}_1(\boldsymbol{\theta}), \dots, \hat{p}_n(\boldsymbol{\theta}))'$ be the maximizer of $\ell_{\text{SEL}}(\mathbf{p})$ under the normalization constraint (2.3) and the parameter constraints (3.4) for a fixed $\boldsymbol{\theta}$. It follows from standard empirical likelihood method that $\hat{p}_i(\boldsymbol{\theta}) = n^{-1} [1 + \boldsymbol{\lambda}' \{w_i \mathbf{g}_i(\boldsymbol{\theta})\}]^{-1}$ for $i \in \mathcal{S}$, where the Lagrange multiplier $\boldsymbol{\lambda} = \boldsymbol{\lambda}(\boldsymbol{\theta})$ is the solution to

$$\mathbf{g}_{\text{SEL}}(\boldsymbol{\lambda}) = \frac{1}{n} \sum_{i \in \mathcal{S}} \frac{w_i \mathbf{g}_i(\boldsymbol{\theta})}{1 + \boldsymbol{\lambda}' \{w_i \mathbf{g}_i(\boldsymbol{\theta})\}} = \mathbf{0}. \quad (3.5)$$

The empirical log-likelihood ratio statistic for $\boldsymbol{\theta}$ under the current setting is given by

$$r_{\text{SEL}}(\boldsymbol{\theta}) = \ell_{\text{SEL}}\{\hat{\mathbf{p}}(\boldsymbol{\theta})\} - \ell_{\text{SEL}}(\hat{\mathbf{p}}) = \sum_{i \in \mathcal{S}} \log\{n \hat{p}_i(\boldsymbol{\theta})\} = - \sum_{i \in \mathcal{S}} \log\{1 + \boldsymbol{\lambda}' w_i \mathbf{g}_i(\boldsymbol{\theta})\}.$$

Let $\hat{\boldsymbol{\theta}}_{\text{SEL}} = \arg \max_{\boldsymbol{\theta} \in \Theta} r_{\text{SEL}}(\boldsymbol{\theta})$ be the maximum sample empirical likelihood estimator of $\boldsymbol{\theta}_N$. Theorem 4 gives the asymptotic distribution of $\hat{\boldsymbol{\theta}}_{\text{SEL}}$ where W_2 is defined in Assumption 5.

Theorem 4. *Suppose that Assumptions 1, 3, 4 and 5 hold. Then*

$$n^{1/2}(\hat{\boldsymbol{\theta}}_{\text{SEL}} - \boldsymbol{\theta}_N) \mid \mathcal{F}_N \xrightarrow{\mathcal{L}} N(0, V_2),$$

where $V_2 = \Sigma_2 \Gamma' W_2^{-1} \Omega W_2^{-1} \Gamma \Sigma_2$ with $\Sigma_2 = (\Gamma' W_2^{-1} \Gamma)^{-1}$.

The result presented in Theorem 4 under the sample empirical likelihood is similar to those in Theorem 1 for the pseudo empirical likelihood, with the crucial differences in defining W_1 for Theorem 1 and W_2 in Theorem 4. For the special case $r = p$, the estimator $\hat{\boldsymbol{\theta}}_{\text{SEL}}$ is attained as the global maximum point with $\hat{p}_i = n^{-1}$ and is the solution to $\sum_{i \in \mathcal{S}} w_i \mathbf{g}_i(\boldsymbol{\theta}) = \mathbf{0}$, which coincides with the survey weighted estimating equations estimator.

Corollary 3. *Suppose that the assumptions of Theorem 4 hold. If $r = p$, then the asymptotic variance-covariance matrix V_2 for $\hat{\boldsymbol{\theta}}_{\text{SEL}}$ reduces to $V_2 = \Gamma^{-1} \Omega (\Gamma')^{-1}$.*

The sample empirical log-likelihood ratio statistic for testing $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_N$ is defined as

$$\text{LR}_{\text{SEL}}(\boldsymbol{\theta}) = 2\{r_{\text{SEL}}(\hat{\boldsymbol{\theta}}_{\text{SEL}}) - r_{\text{SEL}}(\boldsymbol{\theta})\} = 2\{\ell_{\text{SEL}}(\hat{\boldsymbol{\theta}}_{\text{SEL}}) - \ell_{\text{SEL}}(\boldsymbol{\theta})\}$$

for the given $\boldsymbol{\theta}$. We have the following results parallel to Theorem 2 and Corollary 2. Once again, the differences are between W_1 and W_2 involved in the asymptotic distributions.

Theorem 5. *Suppose that Assumptions 1, 3, 4 and 5 hold. Then*

$$\text{LR}_{\text{SEL}}(\boldsymbol{\theta}_N) \mid \mathcal{F}_N \xrightarrow{\mathcal{L}} Q' \Delta_2 Q,$$

where $Q \sim N(0, I_r)$ and $\Delta_2 = \Omega^{1/2} W_2^{-1} \Gamma \Sigma_2 \Gamma' W_2^{-1} \Omega^{1/2}$ with $\Sigma_2 = (\Gamma' W_2^{-1} \Gamma)^{-1}$.

Corollary 4. *Suppose that the assumptions of Theorem 5 hold. If $r = p$, then $\Delta_2 = \Omega^{1/2} W_2^{-1} \Omega^{1/2}$. In particular, if $r = p = 1$, we have $\text{LR}_{\text{SEL}}(\boldsymbol{\theta}_N) \xrightarrow{\mathcal{L}} (\Omega/W_2) \chi^2(1)$.*

For a general linear or nonlinear hypothesis $H_0: R(\boldsymbol{\theta}_N) = \mathbf{0}$ versus $H_1: R(\boldsymbol{\theta}_N) \neq \mathbf{0}$, the restricted maximum sample empirical likelihood estimator of $\boldsymbol{\theta}$ under H_0 is defined as $\hat{\boldsymbol{\theta}}_{\text{SEL}}^R = \arg \max_{\boldsymbol{\theta} \in \Theta^R} r_{\text{SEL}}(\boldsymbol{\theta})$, where $\Theta^R = \{\boldsymbol{\theta} \mid \boldsymbol{\theta} \in \Theta \text{ and } R(\boldsymbol{\theta}) = \mathbf{0}\}$. The sample empirical log-likelihood ratio statistic for testing H_0 against H_1 is given by

$$\text{LR}_{\text{SEL}}(\boldsymbol{\theta}_N \mid H_0) = 2\{r_{\text{SEL}}(\hat{\boldsymbol{\theta}}_{\text{SEL}}) - r_{\text{SEL}}(\hat{\boldsymbol{\theta}}_{\text{SEL}}^R)\} = 2\{\ell_{\text{SEL}}(\hat{\boldsymbol{\theta}}_{\text{SEL}}) - \ell_{\text{SEL}}(\hat{\boldsymbol{\theta}}_{\text{SEL}}^R)\}.$$

Theorem 6. *Suppose that the assumptions of Theorem 3 hold. If the function $R(\boldsymbol{\theta})$ is twice continuously differentiable and $\Phi(\boldsymbol{\theta}_N) = \partial R(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} \mid_{\boldsymbol{\theta} = \boldsymbol{\theta}_N}$ has rank k , then*

$$\text{LR}_{\text{SEL}}(\boldsymbol{\theta}_N \mid H_0) \mid \mathcal{F}_N \xrightarrow{\mathcal{L}} Q' \Delta_2^R Q,$$

where $Q \sim N(0, I_r)$, $\Delta_2^R = \Omega^{1/2} W_2^{-1} \Gamma \Sigma_2 \Phi' (\Phi \Sigma_2 \Phi')^{-1} \Phi \Sigma_2 \Gamma' W_2^{-1} \Omega^{1/2}$, and $\Phi = \Phi(\boldsymbol{\theta}_N)$.

The term W_2 for the sample empirical likelihood is different from W_1 for the pseudo empirical likelihood and can be estimated by

$$\hat{W}_2 = nN^{-2} \sum_{i \in \mathcal{S}} w_i^2 \mathbf{g}_i(\hat{\boldsymbol{\theta}}_{\text{SEL}}) \mathbf{g}_i(\hat{\boldsymbol{\theta}}_{\text{SEL}})'$$

The other two component Γ and Ω can be respectively estimated by $\hat{\Gamma}_{\text{SEL}} = N^{-1} \sum_{i \in \mathcal{S}} w_i \partial \mathbf{g}_i(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} \mid_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_{\text{SEL}}}$ and $\hat{\Omega}_{\text{SEL}} = nN^{-2} v\{\hat{U}_n(\hat{\boldsymbol{\theta}}_{\text{SEL}})\}$, where $v\{\hat{U}_n(\cdot)\}$ is given in Assumption 2.

3.4. Design-based variable selection

Public-use survey data may contain observations on many variables. Variable selection is a useful technique when fitting a statistical model involving many covariates. The pseudo empirical likelihood and the sample empirical likelihood provide design-based approaches to variable selection through a penalized pseudo or sample empirical likelihood method.

Suppose that $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)'$ and $p_{\tau_n}(\cdot)$ is a pre-specified penalty function with regularization parameter τ_n . The penalized pseudo empirical likelihood (PPEL) function of $\boldsymbol{\theta}$ is defined as

$$l_{\text{PPEL}}(\boldsymbol{\theta}) = -n \sum_{i \in \mathcal{S}} \tilde{w}_i(\mathcal{S}) \log\{1 + \boldsymbol{\lambda}' \mathbf{g}_i(\boldsymbol{\theta})\} - n \sum_{j=1}^p p_{\tau_n}(|\theta_j|),$$

where the Lagrange multiplier $\boldsymbol{\lambda}$ solves $\mathbf{g}_{\text{PEL}}(\boldsymbol{\lambda}) = \mathbf{0}$ given by (3.3). The penalized sample empirical likelihood (PSEL) function is defined as

$$l_{\text{PSEL}}(\boldsymbol{\theta}) = - \sum_{i \in \mathcal{S}} \log\{1 + \boldsymbol{\lambda}' w_i \mathbf{g}_i(\boldsymbol{\theta})\} - n \sum_{j=1}^p p_{\tau_n}(|\theta_j|),$$

where the Lagrange multiplier $\boldsymbol{\lambda}$ solves $\mathbf{g}_{\text{SEL}}(\boldsymbol{\lambda}) = \mathbf{0}$ given by (3.5).

The tuning parameter τ_n for the penalized pseudo empirical likelihood or the penalized sample empirical likelihood needs to be appropriately selected by a data-driven method. Various techniques have been proposed in the literature, including the generalized cross-validation method and the BIC method. Further details can be found in Fan and Li [10] and Wang et al. [25].

Let $\boldsymbol{\theta}_N = (\theta_{N_1}, \dots, \theta_{N_p})'$ be defined by (2.1). The maximum penalized pseudo empirical likelihood estimator of $\boldsymbol{\theta}_N$ is defined as $\hat{\boldsymbol{\theta}}_{\text{PPEL}} = \arg \max_{\boldsymbol{\theta}} l_{\text{PPEL}}(\boldsymbol{\theta})$ and the maximum penalized sample empirical likelihood estimator of $\boldsymbol{\theta}_N$ is defined as $\hat{\boldsymbol{\theta}}_{\text{PSEL}} = \arg \max_{\boldsymbol{\theta}} l_{\text{PSEL}}(\boldsymbol{\theta})$. Both estimators enjoy the design-based oracle property for variable selection in the sense that $\Pr(\hat{\theta}_{N_j} = 0 \mid \mathcal{F}_N) \rightarrow 1$ as $N \rightarrow \infty$ if $\theta_{N_j} = 0$, where $\hat{\theta}_{N_j}$ is the corresponding component of $\hat{\boldsymbol{\theta}}_{\text{PPEL}}$ or $\hat{\boldsymbol{\theta}}_{\text{PSEL}}$ for estimating θ_{N_j} .

4. Bootstrap calibrated empirical likelihood methods

One of the most crucial features of public-use survey data files is the inclusion of replication weights. The guiding principle for the creation of replication weights is that they provide valid results on variance estimation, as outlined in Assumption 2. The major results presented in Section 3 involve the estimation of the design-based variance Ω using the replication weights, and inferential procedures are developed based on the limiting distributions presented in the theorems and corollaries.

A promising approach for practical implementations of the EL-based tests is the bootstrap calibration method. The asymptotic distributions are approximated by the empirical distribution of the replicate copies of the empirical likelihood ratio statistic using the bootstrap weights. However, theoretical justifications of the bootstrap calibration method can be a challenging task and need to be developed case-by-case. In this section, we describe a bootstrap procedure for scenarios where the survey design is single-stage PPS sampling with small sampling fractions and the final survey weights are the calibration weights. The auxiliary variables with known population totals used for calibration are available to the producers of the data files but unavailable to data users. Theoretical justifications of the procedure are given in the Appendix A.

The set of variables used for calibration weighting, denoted by \mathbf{z} , often differs from the set of covariates (\mathbf{x}) appearing in the final dataset. Let $T_{\mathbf{z}}$ be the known population totals for auxiliary variables \mathbf{z} . Let $d_i = 1/\pi_i$ be the original design weights and let $\{(y_i, \mathbf{x}_i, \mathbf{z}_i, d_i), i \in \mathcal{S}\}$ be the preliminary survey dataset for weight construction. The calibration weights w_i are obtained by minimizing a

distance measure $D(\mathbf{w}, \mathbf{d})$ between $\mathbf{w} = (w_1, \dots, w_n)'$ and $\mathbf{d} = (d_1, \dots, d_n)'$ subject to the calibration constraints $\sum_{i \in \mathcal{S}} w_i \mathbf{z}_i = T_{\mathbf{z}}$. There are different distance measures available for calibration weighting. Wu and Lu [27] contains an overview on computational algorithms and finite sample behaviours of weights from alternative calibration weighting methods. We consider the simple chisquare distance $D(\mathbf{w}, \mathbf{d}) = \sum_{i \in \mathcal{S}} (w_i - d_i)^2 / d_i$, which leads to closed form expressions for the final calibrated weights w_i . Let $\{(y_i, \mathbf{x}_i, \mathbf{z}_i, w_i, d_i), i \in \mathcal{S}\}$ be the transitional survey dataset under construction without replication weights.

We present bootstrap procedures for the sample empirical likelihood method on testing $H_0: \boldsymbol{\theta}_N = \boldsymbol{\theta}_{N_0}$ against $H_1: \boldsymbol{\theta}_N \neq \boldsymbol{\theta}_{N_0}$. The procedures are also valid for the pseudo empirical likelihood method. The proposed bootstrap procedures consist of the following steps.

1. Select a bootstrap sample \mathcal{S}^* of size n from the original sample \mathcal{S} using simple random sampling with replacement. Denote the bootstrap sample data by $\{(y_i, \mathbf{x}_i, \mathbf{z}_i, w_i, d_i), i \in \mathcal{S}^*\}$. Note that \mathcal{S}^* may contain duplicated units from \mathcal{S} .

2. Compute the set of bootstrap weights $\{w_i^*, i \in \mathcal{S}^*\}$ by minimizing the distance measure $D(\mathbf{w}^*, \mathbf{d}) = \sum_{i \in \mathcal{S}^*} (w_i^* - d_i)^2 / d_i$ subject to the bootstrap version of the calibration constraints $\sum_{i \in \mathcal{S}^*} w_i^* \mathbf{z}_i = T_{\mathbf{z}}$. Note that $T_{\mathbf{z}} = \sum_{i \in \mathcal{S}} w_i \mathbf{z}_i$.

3. Define the bootstrap version of the sample empirical likelihood ratio function $r_{\text{SEL}}(\boldsymbol{\theta})$ as

$$r_{\text{SEL}}^*(\boldsymbol{\theta}) = - \sum_{i \in \mathcal{S}^*} \log\{1 + \boldsymbol{\lambda}' w_i^* \mathbf{g}_i(\boldsymbol{\theta})\},$$

where $\boldsymbol{\lambda}$ is the solution to $g_{\text{SEL}}^*(\boldsymbol{\lambda}) = n^{-1} \sum_{i \in \mathcal{S}^*} \{w_i^* \mathbf{g}_i(\boldsymbol{\theta})\} / [1 + \boldsymbol{\lambda}' \{w_i^* \mathbf{g}_i(\boldsymbol{\theta})\}] = \mathbf{0}$. Compute the bootstrap version of the estimator $\hat{\boldsymbol{\theta}}_{\text{SEL}}^* = \arg \max_{\boldsymbol{\theta} \in \Theta} r_{\text{SEL}}^*(\boldsymbol{\theta})$ and the bootstrap version of the SEL ratio statistic $\text{LR}_{\text{SEL}}^*(\hat{\boldsymbol{\theta}}_{\text{SEL}}) = 2\{r_{\text{SEL}}^*(\hat{\boldsymbol{\theta}}_{\text{SEL}}) - r_{\text{SEL}}^*(\hat{\boldsymbol{\theta}}_{\text{SEL}})\}$, where $\hat{\boldsymbol{\theta}}_{\text{SEL}}$ is the estimator obtained from the original survey dataset $\{(y_i, \mathbf{x}_i, w_i), i \in \mathcal{S}\}$.

4. Repeat Steps 1-3 a large number B times, independently, to obtain B values of the bootstrap version of the SEL ratio statistic as $\{\text{LR}_{\text{SEL}}^{*(1)}(\hat{\boldsymbol{\theta}}_{\text{SEL}}), \dots, \text{LR}_{\text{SEL}}^{*(B)}(\hat{\boldsymbol{\theta}}_{\text{SEL}})\}$.

Let b_α be the upper α quantile from the empirical distribution of the values of the bootstrap version $\{\text{LR}_{\text{SEL}}^{*(1)}(\hat{\boldsymbol{\theta}}_{\text{SEL}}), \dots, \text{LR}_{\text{SEL}}^{*(B)}(\hat{\boldsymbol{\theta}}_{\text{SEL}})\}$. The α -level SEL ratio test rejects $H_0: \boldsymbol{\theta}_N = \boldsymbol{\theta}_{N_0}$ if $\text{LR}_{\text{SEL}}(\boldsymbol{\theta}_N) > b_\alpha$. The bootstrap calibrated $1 - \alpha$ level confidence region for $\boldsymbol{\theta}_N$ is given by $\mathcal{C}_{\text{BT}} = \{\boldsymbol{\theta} \mid \text{LR}_{\text{SEL}}(\boldsymbol{\theta}) \leq b_\alpha\}$. It is shown in the Appendix that this confidence region has correct asymptotic coverage probability.

The bootstrap procedures described above can be implemented through additional columns of replication weights to produce a public-use data file. Let $\{w_i^*, i \in \mathcal{S}^*\}$ be a set of bootstrap weights described in Step 2. Let h_i be the number of times that unit $i \in \mathcal{S}$ is selected in \mathcal{S}^* . Note that $0 \leq h_i \leq n$ and $\sum_{i \in \mathcal{S}} h_i = n$. The b th set of replication weights are constructed as $\{w_i^{(b)} = h_i w_i^*, i \in \mathcal{S}\}$. Repeat the process for $b = 1, \dots, B$, independently, to create B

sets of replication weights. The bootstrap version $\text{LR}_{\text{SEL}}^*(\hat{\boldsymbol{\theta}}_{\text{SEL}})$ of the SEL ratio statistic can be computed by using the (\mathbf{x}, y) from the data file in conjunction with the set of replication weights.

5. Simulation studies

In this section we report results from simulation studies on the finite sample performances of our proposed methods. The finite population $\{(y_i, x_{i1}, x_{i2}, x_{i3}), i = 1, 2, \dots, N\}$ with size N was generated from the following super population model

$$y_i = \mathbf{x}'_i \boldsymbol{\theta} + \sigma \varepsilon_i, \quad i = 1, 2, \dots, N,$$

where $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2, \theta_3)' = (1, 1, 1, 1)'$, $\mathbf{x}_i = (1, x_{i1}, x_{i2}, x_{i3})'$, $x_{i1} \sim \text{Bernoulli}(0.5)$, $x_{i2} \sim \text{Uniform}(0, 1)$, $x_{i3} \sim 0.5 + \text{Exponential}(2)$, and the ε_i 's are iid $N(0, 1)$. We consider three cases for the variance σ^2 of the error terms: (i) $\sigma = \sigma_1 = 1$; (ii) $\sigma = \sigma_2 = 3$; and (iii) $\sigma = \sigma_3 = [\text{Var}(\eta)(1/\rho^2 - 1)]^{1/2}$ with $\eta = \mathbf{x}'\boldsymbol{\theta}$ and $\rho = 0.8$. This is the controlled correlation coefficient between y and the linear predictor η .

The finite population parameters $\boldsymbol{\theta}_N = (\theta_{N0}, \theta_{N1}, \theta_{N2}, \theta_{N3})'$ under the linear regression model are defined as the solution to the census estimating equations $\sum_{i=1}^N \mathbf{g}(\mathbf{x}_i, y_i, \boldsymbol{\theta}_N) = \mathbf{0}$, where $\mathbf{g}(\mathbf{x}, y, \boldsymbol{\theta}) = \mathbf{x}(y - \mathbf{x}'\boldsymbol{\theta})$. With a large N , the values of $\boldsymbol{\theta}_N$ are almost identical to the model parameters for the superpopulation. Our simulation studies focus on examining the size and power of the proposed pseudo and sample empirical likelihood ratio tests. We consider α -level tests for two hypotheses: (1) $H_0: \theta_{N1} = 1.0$ versus $H_1: \theta_{N1} = b$; and (2) $H_0: \theta_{N1} - \theta_{N2} = 0$ versus $H_1: (\theta_{N1}, \theta_{N2}) = (b_1, b_2)$, for selected values of b and (b_1, b_2) , with $\alpha = 0.05$ for both cases.

In survey practice, the process of creating the final survey weights w_i and the final replication weights $w_i^{(b)}$, $b = 1, 2, \dots, B$ can be very complicated. It depends on the original survey design, the scenarios for nonresponse, and the amount of known auxiliary information for calibration weighting. The replication weights often involve ad hoc approximations since many complex survey designs do not have precise bootstrap procedures or other resampling methods to produce final replication weights for general inferences. Rao and Wu [22] and Rao, Wu and Yue [23] contain further details on the topic. To make repeated simulation runs feasible, we consider single stage unequal probability sampling for the initial survey design, with the inclusion probabilities π_i proportional to x_{i3} . The final survey weights and the final replication weights are created under two scenarios:

A. The final survey weights are calibrated over the known population totals of the x_1 and x_2 variables but unit nonresponse is not involved.

B. The final survey weights are adjusted for uniform unit nonresponse and calibrated over the known population totals of the x_1 and x_2 variables.

For each of the two scenarios, there are two major tasks for each simulated sample: compute the final survey weights w_i and create valid final replication weights $w_i^{(b)}$, $i \in \mathcal{S}$. For single stage PPS sampling without replacement with

a negligible sampling fraction, the with-replacement bootstrap procedures described in Section 4 produce final replication weights that satisfy Assumption 2 and are also valid for the bootstrap calibration method described in Section 4. Let \mathcal{S}_0 be the set of initial sampled units and n_0 be the initial sample size under the original survey design and let \mathcal{S} be the set of units included in the final sample and n be the final sample size.

Under Scenario A, we have $\mathcal{S} = \mathcal{S}_0$ and $n = n_0$ in the absence of unit nonresponse. The final weights are calibrated over the known population totals of x_1 and x_2 . The replication weights are created based on the method described in Section 4. Under Scenario B, let $d_i = 1/\pi_i$ be the initial design weights, $i \in \mathcal{S}_0$. With uniform unit nonresponse, each unit in \mathcal{S}_0 has a constant probability to be a respondent, and the final set \mathcal{S} of respondents has a random sample size. The unit nonresponse adjusted survey weights are computed as

$$d_{0i} = d_i \left(\sum_{k \in \mathcal{S}_0} d_k \right) / \left(\sum_{j \in \mathcal{S}} d_j \right), \quad i \in \mathcal{S}.$$

This is the so-called ratio adjustment for uniform unit nonresponse and the adjusted survey weights satisfy $\sum_{i \in \mathcal{S}} d_{0i} = \sum_{j \in \mathcal{S}_0} d_j$. Treating the set of adjusted weights $\{d_{0i}, i \in \mathcal{S}\}$ as the “original” design weights, the final survey weights and replication weights under the calibration constraints are created by following the same procedures used in Scenario A.

Simulation samples of size $n = 400$ are selected for Scenario A from the population by the randomized systematic PPS sampling method (Goodman and Kish, [11]; Hartley and Rao, [12]). For Scenario B, initial samples of size $n_0 = 571$ are selected by the same PPS sampling method. The unit response probabilities are set to be uniform at 0.7, resulting in final samples with expected sample size $E(n) = 400$. For both scenarios, we choose the finite population sizes as $N = 20,000$ and $4,000$ such that the sampling fractions are $n/N = 2\%$ and 10% , the first case represents negligible sampling fractions and the second case is for non-negligible sampling fractions. The final survey weights and the $B = 500$ sets of final replication weights are created for the given scenario.

TABLE 1
Size of the PEL and SEL ratio tests assuming standard χ^2 limiting distributions

n/N		PEL			SEL		
		σ_1	σ_2	σ_3	σ_1	σ_2	σ_3
$H_0: \theta_{N_1} = 1.0$ versus $H_1: \theta_{N_1} = b$							
2%	$b = 1.0$	0.191	0.167	0.189	0.167	0.141	0.167
10%	$b = 1.0$	0.194	0.170	0.193	0.164	0.141	0.164
$H_0: \theta_{N_1} = \theta_{N_2}$ versus $H_1: (\theta_{N_1}, \theta_{N_2}) = (b_1, b_2)$							
2%	$b_1 = b_2 = 1.0$	0.262	0.227	0.260	0.204	0.186	0.206
10%	$b_1 = b_2 = 1.0$	0.261	0.248	0.264	0.205	0.197	0.210

We compute the power of the PEL and SEL ratio tests for $H_0: \theta_{N_1} = 1.0$ versus $H_1: \theta_{N_1} = b$ and for $H_0: \theta_{N_1} = \theta_{N_2}$ versus $H_1: (\theta_{N_1}, \theta_{N_2}) = (b_1, b_2)$ for selected values of b and (b_1, b_2) . The power for $b = 1.0$ and $(b_1, b_2) = (1.0, 1.0)$ represents the size of the test, which is set at the level 0.05. Results are based on 2,000 simulation runs. As a warning message for possible misuse of the PEL and SEL

TABLE 2
 Power of the PEL tests for $H_0: \theta_{N_1} = 1.0$ versus $H_1: \theta_{N_1} = b$ when $n/N = 2\%$

		$b =$	0.50	0.75	1.00	1.25	1.50
A	I	σ_1	0.991	0.647	0.046	0.637	0.994
		σ_2	0.385	0.142	0.056	0.102	0.316
		σ_3	0.833	0.330	0.054	0.304	0.840
	II	σ_1	0.992	0.637	0.053	0.633	0.993
		σ_2	0.340	0.118	0.051	0.124	0.332
		σ_3	0.834	0.345	0.050	0.302	0.841
	III	σ_1	0.989	0.656	0.056	0.635	0.996
		σ_2	0.359	0.123	0.057	0.112	0.316
		σ_3	0.836	0.320	0.054	0.300	0.840
	IV	σ_1	0.986	0.654	0.050	0.638	0.996
		σ_2	0.352	0.152	0.052	0.104	0.324
		σ_3	0.818	0.314	0.044	0.254	0.828
	V	σ_1	0.992	0.635	0.053	0.626	0.993
		σ_2	0.334	0.115	0.048	0.123	0.328
		σ_3	0.831	0.343	0.049	0.300	0.834
B	I	σ_1	0.913	0.395	0.050	0.373	0.921
		σ_2	0.226	0.094	0.040	0.087	0.181
		σ_3	0.613	0.207	0.048	0.186	0.576
	II	σ_1	0.916	0.403	0.055	0.378	0.926
		σ_2	0.199	0.079	0.038	0.078	0.191
		σ_3	0.581	0.207	0.052	0.173	0.574
	III	σ_1	0.923	0.413	0.049	0.380	0.913
		σ_2	0.202	0.099	0.053	0.083	0.194
		σ_3	0.577	0.214	0.042	0.193	0.569
	V	σ_1	0.921	0.416	0.059	0.397	0.929
		σ_2	0.207	0.084	0.045	0.083	0.203
		σ_3	0.601	0.213	0.056	0.186	0.593

based tests, we first show that naively assuming the limiting distributions of the PEL and the SEL ratio tests with public survey data files as standard chisquares leads to invalid results. The sizes of the tests under Scenario A with different settings are presented in Table 1. It is apparent from Table 1 that the test sizes are off by a large margin relative to the nominal value 0.05 for all cases ranging from 0.141 to 0.194 for the first test and 0.186 to 0.264 for the second test.

The limiting distributions of the PEL and the SEL ratio tests generally follow the distribution of a quadratic form presented in Section 3. We consider four methods to determine the critical region for each test: I. Monte Carlo approximations to the distribution of the quadratic form using the estimated eigenvalues and the weighted χ^2 distribution; II. The first-order Rao-Scott correction method; III. The second-order Rao-Scott correction method; IV. The Bootstrap calibration method as described in Section 4. We also included a fifth method for comparisons: V. The Wald-test based on the point estimator $\hat{\theta}$ and the variance estimator $v(\hat{\theta})$ for $\theta = \theta_{N_1}$ or $\theta = \theta_{N_1} - \theta_{N_2}$ using standard normal approximation to $(\hat{\theta} - \theta)/\{v(\hat{\theta})\}^{1/2}$. Method I uses the limiting distributions presented in Section 3. Methods I, II and III all require the estimation of eigenvalues of the matrix $\Delta_1, \Delta_1^R, \Delta_2$ or Δ_2^R . The bootstrap calibration method IV is extremely time consuming for repeated simulations and the results are only included for Scenario A with 500 simulation runs.

TABLE 3
 Power of the SEL tests for $H_0: \theta_{N_1} = 1.0$ versus $H_1: \theta_{N_1} = b$ when $n/N = 2\%$

		$b =$	0.50	0.75	1.00	1.25	1.50
A	I	σ_1	0.995	0.678	0.058	0.664	0.995
		σ_2	0.393	0.150	0.059	0.116	0.331
		σ_3	0.851	0.345	0.062	0.326	0.853
	II	σ_1	0.994	0.674	0.062	0.665	0.994
		σ_2	0.353	0.130	0.058	0.140	0.353
		σ_3	0.857	0.362	0.057	0.336	0.857
	III	σ_1	0.995	0.667	0.059	0.664	0.996
		σ_2	0.353	0.131	0.070	0.140	0.352
		σ_3	0.848	0.346	0.066	0.342	0.845
	IV	σ_1	0.986	0.652	0.050	0.634	0.996
		σ_2	0.344	0.150	0.046	0.100	0.316
		σ_3	0.816	0.308	0.040	0.248	0.818
	V	σ_1	0.992	0.635	0.053	0.626	0.993
		σ_2	0.334	0.115	0.048	0.123	0.328
		σ_3	0.831	0.343	0.049	0.300	0.834
B	I	σ_1	0.943	0.469	0.076	0.447	0.938
		σ_2	0.264	0.118	0.057	0.115	0.218
		σ_3	0.662	0.245	0.066	0.236	0.630
	II	σ_1	0.939	0.472	0.075	0.448	0.942
		σ_2	0.228	0.101	0.059	0.105	0.236
		σ_3	0.647	0.241	0.073	0.225	0.632
	III	σ_1	0.940	0.490	0.071	0.454	0.936
		σ_2	0.247	0.128	0.068	0.110	0.238
		σ_3	0.651	0.232	0.063	0.227	0.631
	V	σ_1	0.921	0.416	0.059	0.397	0.929
		σ_2	0.207	0.084	0.045	0.083	0.203
		σ_3	0.601	0.213	0.056	0.186	0.593

Tables 2 and 3 summarize the results on the size and power of the tests for $H_0: \theta_{N_1} = 1.0$ versus $H_1: \theta_{N_1} = b$ for PEL and SEL, respectively, with $n/N = 2\%$. The results for $b = 1.0$ correspond to the size of the test with nominal value 0.05 and the results for $b \neq 1.0$ represent the actual power of the test. Tables 4 and 5 summarize the results on the size and power of the tests for $H_0: \theta_{N_1} = \theta_{N_2}$ versus $H_1: (\theta_{N_1}, \theta_{N_2}) = (b_1, b_2)$. The results for $(b_1, b_2) = (1.0, 1.0)$ correspond to the size of the test and the results for other values of (b_1, b_2) represent the power of the test. Simulation results corresponding to $n/N = 10\%$ are reported in the online version of the paper (Zhao et al., [31]).

Major observations of the simulation results in Tables 2–5 can be summarized as follows. (1) All three approaches (i.e., PEL, SEL and Wald) have test sizes close to the nominal value 0.05 for almost all cases. The PEL based tests perform the best in terms of valid test size while the SEL based tests have a few cases with sizes bigger than 0.065. (2) The tests are generally more powerful when the error variance σ^2 is smaller (the cases with σ_1 and σ_2), where the auxiliary variables used for calibration weighting have stronger correlation to the response variable. (3) Both the first and the second order Rao-Scott corrections (entries under II and III) provide similar results compared to the ones using the actual limiting distributions (entries under I). (4) The validity of the replication weights is justified for cases with small sampling fractions but the results based on the

TABLE 4
 Power of the PEL tests for $H_0: \theta_{N_1} = \theta_{N_2}$ versus $H_1: (\theta_{N_1}, \theta_{N_2}) = (b_1, b_2)$ when $n/N = 2\%$

		$(b_1, b_2) =$	(1.0, 2.0)	(1.0, 1.5)	(1.0, 1.0)	(1.5, 1.0)	(2.0, 1.0)
A	I	σ_1	0.997	0.658	0.056	0.987	1.000
		σ_2	0.357	0.117	0.041	0.248	0.784
		σ_3	0.866	0.306	0.055	0.754	1.000
	II	σ_1	0.998	0.672	0.054	0.985	1.000
		σ_2	0.339	0.129	0.053	0.263	0.779
		σ_3	0.876	0.312	0.055	0.762	1.000
	III	σ_1	0.997	0.642	0.055	0.988	1.000
		σ_2	0.339	0.113	0.062	0.255	0.779
		σ_3	0.862	0.321	0.056	0.737	1.000
	IV	σ_1	0.998	0.638	0.048	0.992	1.000
		σ_2	0.318	0.118	0.054	0.242	0.816
		σ_3	0.822	0.272	0.040	0.742	1.000
	V	σ_1	0.998	0.724	0.059	0.558	0.991
		σ_2	0.364	0.148	0.050	0.110	0.320
		σ_3	0.867	0.370	0.058	0.294	0.823
B	I	σ_1	0.931	0.381	0.051	0.864	1.000
		σ_2	0.169	0.079	0.055	0.145	0.506
		σ_3	0.561	0.179	0.051	0.444	0.974
	II	σ_1	0.937	0.408	0.054	0.863	0.999
		σ_2	0.182	0.080	0.039	0.143	0.465
		σ_3	0.599	0.177	0.058	0.461	0.980
	III	σ_1	0.937	0.392	0.052	0.855	1.000
		σ_2	0.194	0.067	0.046	0.134	0.516
		σ_3	0.593	0.177	0.053	0.451	0.981
	V	σ_1	0.941	0.486	0.050	0.342	0.896
		σ_2	0.236	0.101	0.049	0.090	0.187
		σ_3	0.638	0.232	0.053	0.174	0.570

estimated eigenvalues (entries under I, II and III) seem to work well even if $n/N = 10\%$. (5) The bootstrap calibration method (entries under IV) works very well for $n/N = 2\%$ for all cases. For cases with the large sampling fraction $n/N = 10\%$, the size of the test for $H_0: \theta_{N_1} = \theta_{N_2}$ with $\sigma = \sigma_1$ is around 0.02 for both PEL and SEL, showing the sensitivity of the replication weights on the bootstrap calibrated tests. (6) The Wald test has similar performance to SEL based tests in some cases but is less powerful in some other cases.

Results on the performance of the empirical likelihood methods for parameters defined through nonsmooth estimating functions are reported in the online version of the paper (Zhao et al., [31]).

6. An application to the GSS 2016 dataset

The General Social Survey (GSS) is an annual cross-sectional survey conducted by Statistics Canada since 1985. The survey gathers data on social trends in order to monitor changes in the living conditions and the well-being of Canadians, and to provide information on specific social policy issues. The 2016 GSS focused on Canadians at Work and Home, and collected information on the lifestyle behaviour of Canadians that affects their health and well-being, both in workplace and home. The survey covered individuals aged 15 years and older living in pri-

TABLE 5
Power of the SEL tests for $H_0: \theta_{N1} = \theta_{N2}$ versus $H_1: (\theta_{N1}, \theta_{N2}) = (b_1, b_2)$ when $n/N = 2\%$

		$(b_1, b_2) =$	(1.0, 2.0)	(1.0, 1.5)	(1.0, 1.0)	(1.5, 1.0)	(2.0, 1.0)
A	I	σ_1	0.997	0.669	0.065	0.988	1.000
		σ_2	0.376	0.135	0.049	0.269	0.798
		σ_3	0.865	0.322	0.064	0.771	1.000
	II	σ_1	0.998	0.687	0.062	0.987	1.000
		σ_2	0.360	0.140	0.061	0.288	0.797
		σ_3	0.878	0.332	0.063	0.778	1.000
	III	σ_1	0.998	0.664	0.063	0.989	1.000
		σ_2	0.361	0.131	0.069	0.278	0.789
		σ_3	0.868	0.337	0.068	0.756	1.000
	IV	σ_1	0.998	0.624	0.040	0.990	1.000
		σ_2	0.294	0.112	0.046	0.232	0.812
		σ_3	0.798	0.252	0.038	0.732	1.000
	V	σ_1	0.998	0.724	0.059	0.558	0.991
		σ_2	0.364	0.148	0.050	0.110	0.320
		σ_3	0.867	0.370	0.058	0.294	0.823
B	I	σ_1	0.941	0.445	0.076	0.897	1.000
		σ_2	0.226	0.112	0.069	0.188	0.546
		σ_3	0.642	0.219	0.069	0.511	0.982
	II	σ_1	0.947	0.470	0.079	0.897	1.000
		σ_2	0.227	0.120	0.057	0.192	0.522
		σ_3	0.645	0.219	0.079	0.526	0.986
	III	σ_1	0.949	0.454	0.075	0.893	1.000
		σ_2	0.245	0.096	0.066	0.178	0.577
		σ_3	0.644	0.234	0.077	0.531	0.986
	V	σ_1	0.941	0.486	0.050	0.342	0.896
		σ_2	0.236	0.101	0.049	0.090	0.187
		σ_3	0.638	0.232	0.053	0.174	0.570

vate households in the 10 provinces of Canada. Public-use GSS micro data files, which include the final survey weights and 500 sets of bootstrap weights, can be accessed through Statistics Canada's Research Data Centre (RDC) or the Data Liberation Initiative (DLI) at major Canadian universities.

We analyzed a subset of the GSS 2016 data file using the pseudo empirical likelihood and the sample empirical likelihood methods developed in this paper. We explored the relationships between the response variable y on job satisfaction and a set of 14 covariates through logistic regression analysis. The y variable is dichotomized from the original 5-point likert scale, i.e., $y = 1$ if either "Very satisfied" or "Satisfied" and $y = 0$ otherwise. The set of covariates includes x_1 : Gender; x_2 : Marital Status; x_3 : Landed Immigrant Status; x_4 : Citizenship Status; x_5 : Number of Weeks Employed – Past 12 Months; x_6 : Number of Weeks Worked at the Job – Past 12 Months; x_7 : Unionized Job or Covered by Contract or Collective Agreement; x_8 : Being Happy When Working Hard; x_9 : Employment Benefits – Workplace Pension Plan; x_{10} : Employment Benefits – Paid Sick Leave; x_{11} : Employment Benefits – Paid Vacation Leave; x_{12} : Unfair Treatment/Discrimination – Past 12 Months; x_{13} : Age Group; x_{14} : Number of Persons Employed at Work Location. The subset of the data file we used, denoted as \mathcal{S} , consists of $n = 1,552$ individuals who had valid responses to all 15 questions described above. Detailed descriptions of those questions are provided

in the Supplementary Material. The final survey weights w_i and the b th set of bootstrap weights $w_i^{(b)}$ are rescaled such that $\sum_{i \in \mathcal{S}} w_i = n$ and $\sum_{i \in \mathcal{S}} w_i^{(b)} = n$, $b = 1, \dots, 500$. Note that the rescaling does not change the validity of the bootstrap weights for variance estimation as specified in Assumption 2.

We considered the logistic regression model on y given $\mathbf{x} = (1, x_1, \dots, x_{14})'$, which models $\Pr(y = 1 \mid \mathbf{x})$ through the logit link function $\text{logit}\{\Pr(y = 1 \mid \mathbf{x})\} = \mathbf{x}'\boldsymbol{\theta}$, where $\text{logit}(p) = \log\{p/(1-p)\}$ and $\boldsymbol{\theta} = (\theta_0, \theta_1, \dots, \theta_{14})'$. It follows that the odds for job satisfaction is given by

$$\frac{\Pr(y = 1 \mid \mathbf{x})}{\Pr(y = 0 \mid \mathbf{x})} = \prod_{j=0}^{14} \exp(x_j \theta_j).$$

The value $\exp(\theta_j)$ represents the odds ratio (OR) for job satisfaction when x_j changes from 0 to 1 given other covariates.

TABLE 6
GSS Data: Point Estimation, Hypothesis Testing and Variable Selection

Covariate	Estimate	SE	OR	P-Value		Variable Selection	
				PEL	SEL	PEL	SEL
1	-0.029	0.716	0.971	0.967	0.962	0.000	0.000
x_1	-0.261	0.211	0.770	0.202	0.165	0.000	0.000
x_2	0.211	0.224	1.234	0.342	0.312	0.000	0.000
x_3	0.091	0.329	1.095	0.779	0.750	0.000	0.000
x_4	-0.250	0.258	0.778	0.319	0.305	0.000	0.000
x_5	0.017	0.014	1.017	0.205	0.159	0.000	0.000
x_6	-0.012	0.011	0.988	0.274	0.261	0.000	0.000
x_7	0.060	0.235	1.061	0.792	0.785	0.000	0.000
x_8	1.258	0.277	3.518	0.000	0.000	2.196	2.157
x_9	0.095	0.268	1.099	0.704	0.693	0.000	0.000
x_{10}	0.590	0.263	1.803	0.019	0.013	0.000	0.000
x_{11}	0.152	0.260	1.164	0.550	0.536	0.000	0.000
x_{12}	-1.422	0.266	0.241	0.000	0.676	0.000	0.000
x_{13}	0.082	0.099	1.085	0.385	0.340	0.000	0.000
x_{14}	0.032	0.059	1.032	0.586	0.566	0.000	0.000

Note: The values 0.000 in the last two columns indicate non-significant factors identified by the variable selection procedure.

The estimating function for defining $\boldsymbol{\theta}$ is given by $\mathbf{g}(\mathbf{x}, y, \boldsymbol{\theta}) = \mathbf{x}\{y - \mu(\mathbf{x}'\boldsymbol{\theta})\}$, where $\mu(\mathbf{x}'\boldsymbol{\theta}) = \exp(\mathbf{x}'\boldsymbol{\theta}) / \{1 + \exp(\mathbf{x}'\boldsymbol{\theta})\}$. Let $\boldsymbol{\theta}_N = (\theta_{N_0}, \theta_{N_1}, \dots, \theta_{N_{14}})'$ be the finite population parameters defined by the census estimating equations. We computed the point estimates, the standard errors (SE), the odds ratios (OR) and the p-values for testing $H_0: \theta_{N_j} = 0$ versus $H_1: \theta_{N_j} \neq 0$, $j = 0, 1, \dots, 14$ using the pseudo empirical likelihood (PEL) and the sample empirical likelihood (SEL) methods. Note that we have $r = p$ in this case and the point estimates, the SE and the OR are the same under the two methods. The p-values for hypothesis tests were computed using the first-order Rao-Scott correction as described at the end of Section 3.2. The SCAD penalty function proposed by Fan and Li [10] was used for variable selection.

Results of estimation, hypothesis testing and variable selection are presented in Table 6. The first major observation is that the pseudo empirical likelihood

and the sample empirical likelihood provide similar results for almost all cases, with only one noticeable exception on the p-value for testing $H_0: \theta_{N12} = 0$. The second observation is that only three covariates, x_8 : *Being Happy When Working Hard*, x_{10} : *Employment Benefits – Paid Sick Leave*, and x_{12} : *Unfair Treatment/Discrimination – Past 12 Months*, show significance to the response variable on job satisfaction from individual tests given all other covariates in the model. The variable selection results, however, point to the fact that x_8 is the most significant factor on job satisfaction.

7. Additional remarks

Public-use survey data files might be utilized by researchers with diverse backgrounds and for different scientific objectives. Descriptive population parameters such as means and proportions, especially at the level of user-defined domains, are often of interest. However, complex survey data have also been used for analytic purposes. One important application is hypothesis tests in the presence of nuisance parameters. Binder and Patak [5] discussed an estimating equation based test on one parameter in the presence of another nuisance parameter. Oguz-Alper and Berger [16] presented a profile empirical likelihood test with nuisance parameters under the setting that detailed design information such as the first order inclusion probabilities and the population auxiliary information used for calibration are available. They showed that the limiting distribution of the empirical likelihood ratio statistic follows a standard chisquare for certain sampling designs. General results, such as Theorems 1–6 presented in Section 3, for public-use survey data are not available in the existing literature. More importantly, naively assuming standard chisquare limiting distributions for the empirical likelihood ratio test statistics for public-use survey data files lead to invalid results, as shown by the simulation results presented in Table 1.

A very important practical problem is variable selection when the survey dataset is used to fit a model involving a large number of covariates. The design-based variable selection techniques described in Section 3.4 are a major contribution of the current paper. Another topic of interest is to test the correctness of the specified model, which is equivalent to testing the unbiasedness of the estimating functions used in the constraints. A pseudo empirical likelihood or a sample empirical likelihood ratio test following Corollary 4 of Qin and Lawless [18] seems to be possible. Detailed procedures are currently under investigation.

The empirical likelihood methods have been an active research topic during the past three decades, with many new developments covering different areas. Rao and Wu [21] contained an overview of empirical likelihood for complex surveys up to 2009. There have been several advances in recent years on empirical likelihood for complex surveys as evidenced by the additional references cited in this paper. Reid [24] provided an overview of likelihood inference in complex settings, and the development of empirical likelihood method for complex survey data received high attention on her list. Our paper addresses a topic with both theoretical and practical importance on analysis of public-use survey data files.

Our proposed methods are valid for any public-use survey data files regardless of the original survey design. However, the bootstrap calibrated tests described in Section 4 put restrictions on how the final replication weights should be produced. Creating final replication weights for valid variance estimation (Assumption 2) has been known to be a challenging task at the data file production stage for complex surveys involving stratification and multi-stage unequal probability sampling. Our simulation results show that constructing replication weights to satisfy the requirements for the bootstrap calibration method is even harder. Another important topic is on how to handle item nonresponse for public-use data files. Single imputation methods are a popular approach among some statistical agencies to produce a single complete data file for public users. How to create replication weights for data files in the presence of imputation for missing values is a topic that deserves high attention in future research.

Appendix A: Theoretical justification of the bootstrap calibration method

The justification of the bootstrap method essentially involves establishing the bootstrap version of Theorem 5 on the sample empirical likelihood. We consider cases where the final survey weights w_i are calibrated over the known population totals of the \mathbf{z} variables using the chi-square distance $D(\mathbf{w}, \mathbf{d})$. The calibrated weights are given by $w_i = d_i\{1 + \mathbf{z}'_i \boldsymbol{\lambda}_c\}$, where $\boldsymbol{\lambda}_c = (\sum_{i \in \mathcal{S}} d_i \mathbf{z}_i \mathbf{z}'_i)^{-1} (T_{\mathbf{z}} - \hat{T}_{\mathbf{zHT}})$, $T_{\mathbf{z}} = \sum_{i=1}^N \mathbf{z}_i$ and $\hat{T}_{\mathbf{zHT}} = \sum_{i \in \mathcal{S}} d_i \mathbf{z}_i$. Let $\mathbf{B}(\boldsymbol{\theta}) = (\sum_{i=1}^N \mathbf{z}_i \mathbf{z}'_i)^{-1} \sum_{i=1}^N \mathbf{z}_i \mathbf{g}_i(\boldsymbol{\theta})'$ and $\hat{\mathbf{B}}(\boldsymbol{\theta}) = (\sum_{i \in \mathcal{S}} d_i \mathbf{z}_i \mathbf{z}'_i)^{-1} \sum_{i \in \mathcal{S}} d_i \mathbf{z}_i \mathbf{g}_i(\boldsymbol{\theta})'$. Under regularity conditions similar to Assumptions 3–5 on the original survey design, we have $\|\hat{\mathbf{B}}(\boldsymbol{\theta}) - \mathbf{B}(\boldsymbol{\theta})\| = O_p(n^{-1/2})$ uniformly for all $\boldsymbol{\theta} \in \Theta$. Consequently, we have the following asymptotic expansion:

$$\begin{aligned} \hat{U}_n(\boldsymbol{\theta}_N) &= \sum_{i \in \mathcal{S}} w_i \mathbf{g}_i(\boldsymbol{\theta}_N) \\ &= \sum_{i \in \mathcal{S}} d_i [\mathbf{g}_i(\boldsymbol{\theta}_N) + \mathbf{g}_i(\boldsymbol{\theta}_N) \mathbf{z}'_i \boldsymbol{\lambda}_c] \\ &= \sum_{i \in \mathcal{S}} d_i \mathbf{g}_i(\boldsymbol{\theta}_N) + \hat{\mathbf{B}}(\boldsymbol{\theta}_N)' (T_{\mathbf{z}} - \hat{T}_{\mathbf{zHT}}) \\ &= \sum_{i \in \mathcal{S}} d_i \mathbf{g}_i(\boldsymbol{\theta}_N) + \mathbf{B}(\boldsymbol{\theta}_N)' (T_{\mathbf{z}} - \hat{T}_{\mathbf{zHT}}) + o_p(Nn^{-1/2}) \\ &= \sum_{i \in \mathcal{S}} d_i [\mathbf{g}_i(\boldsymbol{\theta}_N) - \mathbf{B}(\boldsymbol{\theta}_N)' \mathbf{z}_i] + \mathbf{B}(\boldsymbol{\theta}_N)' T_{\mathbf{z}} + o_p(Nn^{-1/2}). \end{aligned}$$

Let $\hat{U}_n^c(\boldsymbol{\theta}) = \sum_{i \in \mathcal{S}} d_i \mathbf{e}_i(\boldsymbol{\theta}) + \mathbf{B}(\boldsymbol{\theta})' T_{\mathbf{z}}$, where $\mathbf{e}_i(\boldsymbol{\theta}) = \mathbf{g}_i(\boldsymbol{\theta}) - \mathbf{B}(\boldsymbol{\theta})' \mathbf{z}_i$. We have

$$\begin{aligned} \text{LR}_{\text{SEL}}(\boldsymbol{\theta}_N) &= 2\{r_{\text{SEL}}(\hat{\boldsymbol{\theta}}_{\text{SEL}}) - r_{\text{SEL}}(\boldsymbol{\theta}_N)\} \\ &= nN^{-2} \hat{U}_n^c(\boldsymbol{\theta}_N)' W_2^{-1} \Gamma \Sigma_2 \Gamma' W_2^{-1} \hat{U}_n^c(\boldsymbol{\theta}_N) + o_p(1). \end{aligned}$$

The bootstrap weights $\{w_i^*, \in \mathcal{S}^*\}$ are created by the same calibration procedure over \mathbf{z}_i and $T_{\mathbf{z}}$. Using similar arguments for the asymptotic expansion to $\text{LR}_{\text{SEL}}(\hat{\boldsymbol{\theta}}_N)$ and conditional on the original sample, we have a similar expansion to the bootstrap version of the SEL ratio statistic as

$$\begin{aligned} \text{LR}_{\text{SEL}}^*(\hat{\boldsymbol{\theta}}_{\text{SEL}}) &= 2\{r_{\text{SEL}}^*(\hat{\boldsymbol{\theta}}_{\text{SEL}}^*) - r_{\text{SEL}}^*(\hat{\boldsymbol{\theta}}_{\text{SEL}})\} \\ &= nN^{-2}\hat{U}_n^{c*}(\hat{\boldsymbol{\theta}}_{\text{SEL}})'W_2^{-1}\Gamma\Sigma_2\Gamma'W_2^{-1}\hat{U}_n^{c*}(\hat{\boldsymbol{\theta}}_{\text{SEL}}) + o_p(1), \end{aligned}$$

where $\hat{U}_n^{c*}(\boldsymbol{\theta}) = \sum_{i \in \mathcal{S}^*} d_i^* \mathbf{e}_i^*(\boldsymbol{\theta}) + \mathbf{B}(\boldsymbol{\theta})'T_{\mathbf{z}}$, and $\mathbf{e}_i^*(\boldsymbol{\theta}) = \mathbf{g}_i^*(\boldsymbol{\theta}) - \mathbf{B}(\boldsymbol{\theta})'\mathbf{z}_i^*$ for $i \in \mathcal{S}^*$. To justify the proposed bootstrap calibration method, it suffices to show that, as $n \rightarrow \infty$,

$$\text{Var}\left\{\sum_{i \in \mathcal{S}} d_i \mathbf{e}_i(\boldsymbol{\theta}_N) \mid \mathcal{F}_N\right\} / \text{Var}\left\{\sum_{i \in \mathcal{S}^*} d_i^* \mathbf{e}_i^*(\hat{\boldsymbol{\theta}}_{\text{SEL}}) \mid \mathcal{S}\right\} \rightarrow 1, \quad (\text{A.1})$$

where $\text{Var}(\cdot \mid \mathcal{F})$ represents the design-based variance and $\text{Var}(\cdot \mid \mathcal{S})$ denotes the variance under the bootstrap sampling procedure, conditional on the original survey sample \mathcal{S} . The condition (A.1) ensures that both the bootstrap version $\text{LR}_{\text{SEL}}^*(\hat{\boldsymbol{\theta}}_{\text{SEL}})$ and the original version $\text{LR}_{\text{SEL}}(\boldsymbol{\theta}_N)$ have the same asymptotic distribution.

Let $\hat{\boldsymbol{\eta}} = \sum_{i \in \mathcal{S}} d_i \mathbf{e}_i(\boldsymbol{\theta}_N)$ and $\hat{\boldsymbol{\eta}}^* = \sum_{i \in \mathcal{S}^*} d_i^* \mathbf{e}_i^*(\hat{\boldsymbol{\theta}}_{\text{SEL}})$; let $t_i = (nd_i)^{-1}$ and $t_i^* = (nd_i^*)^{-1}$, which play the role of the size variable under PPS sampling. We can rewrite $\hat{\boldsymbol{\eta}}$ and $\hat{\boldsymbol{\eta}}^*$ as $\hat{\boldsymbol{\eta}} = n^{-1} \sum_{i \in \mathcal{S}} \hat{\mathbf{r}}_i$ and $\hat{\boldsymbol{\eta}}^* = n^{-1} \sum_{i \in \mathcal{S}^*} \hat{\mathbf{r}}_i^*$, respectively, where $\hat{\mathbf{r}}_i = \mathbf{e}_i(\boldsymbol{\theta}_N)/t_i$ and $\hat{\mathbf{r}}_i^* = \mathbf{e}_i^*(\hat{\boldsymbol{\theta}}_{\text{SEL}})/t_i^*$. Under the proposed with-replacement bootstrap procedure, we have $\text{Var}(\hat{\boldsymbol{\eta}}^* \mid \mathcal{S}) = \mathbf{S}_{\mathbf{r}}^2/n$, where $\mathbf{S}_{\mathbf{r}}^2 = n^{-1} \sum_{i \in \mathcal{S}} (\mathbf{r}_i - \hat{\boldsymbol{\eta}})(\mathbf{r}_i - \hat{\boldsymbol{\eta}})'$. If the original survey sample is selected by single-stage PPS sampling with replacement method, then $\hat{\boldsymbol{\eta}} = n^{-1} \sum_{i \in \mathcal{S}} \mathbf{g}_i(\boldsymbol{\theta}_N)/t_i$ is the standard Hansen-Hurwitz estimator and the design-based variance $\text{Var}(\hat{\boldsymbol{\eta}} \mid \mathcal{F}_N)$ can be unbiasedly estimated by $n^{-1}\{(n-1)^{-1}\} \sum_{i \in \mathcal{S}} (\mathbf{r}_i - \hat{\boldsymbol{\eta}})(\mathbf{r}_i - \hat{\boldsymbol{\eta}})'$. It follows that $\text{Var}(\hat{\boldsymbol{\eta}} \mid \mathcal{F}_N) / \text{Var}(\hat{\boldsymbol{\eta}}^* \mid \mathcal{S}^*) \rightarrow 1$ as $n \rightarrow \infty$. The result also applies to single-stage PPS sampling without replacement with small sampling fractions.

Acknowledgements

This research was supported by grants from the Natural Sciences and Engineering Research Council (NSERC) of Canada and a Collaborative Research Team Grant from the Canadian Statistical Sciences Institute (CANSSI). Zhao's research was also supported by the scientific research fund for high-level talents of Yunnan University and the National Natural Science Foundation of China (Grant Nos.: 11731011, 11871287).

References

- [1] BERGER, Y. G. and DE LA RIVA TORRES, O. (2012). A unified theory of empirical likelihood ratio confidence intervals for survey data with unequal

- probabilities. *Proceedings of the Survey Research Method Section of the American Statistical Association*, Joint Statistical Meeting, San Diego.
- [2] BERGER, Y. G. and DE LA RIVA TORRES, O. (2014). Empirical likelihood confidence intervals: an application to the EU-SILC household surveys. *Contribution to Sampling Statistics, Contribution to Statistics*: F. Mecatti, P. L. Conti, M. G. Ranalli (editors). Springer, pages 65–84.
 - [3] BERGER, Y. G. and DE LA RIVA TORRES, O. (2016). Empirical likelihood confidence intervals for complex sampling designs. *Journal of Royal Statistical Society, Ser. B*, **78**, 319–341. [MR3454199](#)
 - [4] BINDER, D. A. (1983). The the variances of asymptotically normal estimators from complex surveys. *International Statistical Review*, **51**, 279–292. [MR0731144](#)
 - [5] BINDER, D. A. and PATAK, Z. (1994). Use of estimating functions for estimation from complex surveys. *Journal of the American Statistical Association*, **89**, 1035–1043. [MR1294748](#)
 - [6] CHEN, J. and QIN, J. (1993). Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. *Biometrika*, **80**, 107–116. [MR1225218](#)
 - [7] 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](#)
 - [8] CHEN, J., SITTER, R. R. and WU, C. (2002). Using empirical likelihood methods to obtain range restricted weights in regression estimators for surveys. *Biometrika*, **89**, 230–237. [MR1888365](#)
 - [9] CHEN, S. and KIM, J. K. (2014). Population empirical likelihood for non-parametric inference in survey sampling. *Statistica Sinica*, **24**, 335–355. [MR3183687](#)
 - [10] FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, **96**, 1348–1360. [MR1946581](#)
 - [11] GOODMAN, R. and KISH, L. (1950). Controlled selection – a technique in probability sampling. *Journal of the American Statistical Association*, **45**, 350–372.
 - [12] HARTLEY, H. O. and RAO, J. N. K. (1962). Sampling with unequal probabilities and without replacement. *Annals of Mathematical Statistics*, **33**, 350–374. [MR0143312](#)
 - [13] HARTLEY, H. O. and RAO, J. N. K. (1968). A new estimation theory for sample surveys. *Biometrika*, **55**, 547–557.
 - [14] ISAKI, C. T. and FULLER, W. A. (1982). Survey designs under the regression superpopulation model. *Journal of the American Statistical Association*, **77**, 89–96. [MR0648029](#)
 - [15] KIM, J. K. and WU, C. (2013). Sparse and efficient replication variance estimation for complex surveys. *Survey Methodology*, **39**, 91–120.
 - [16] OGUZ-ALPER, M. and BERGER, Y. G. (2016). Modelling complex survey data with population level information: An empirical likelihood approach. *Biometrika*, **103**, 447–459. [MR3509898](#)

- [17] OWEN, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75**, 237–249. [MR0946049](#)
- [18] QIN, J. and LAWLESS, J. F. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics*, **22**, 300–325. [MR1272085](#)
- [19] RAO, J. N. K. and SOCTT, A. (1981). The analysis of categorical data from complex sample surveys: chi-squared tests for goodness-of-fit and independence in two-way tables. *Journal of the American Statistical Association*, **76**, 221–230. [MR0624328](#)
- [20] RAO, J. N. K. and SCOTT, A. (1984). On chi-squared tests for multi-way tables with cell proportions estimated from survey data. *Annals of Statistics*, **12**, 46–60. [MR0733498](#)
- [21] RAO, J. N. K. and WU, C. (2009). Empirical likelihood methods. *Handbook of Statistics*, Volume 29B, *Sample Surveys: Inference and Analysis*, edited by D. Pfeiffermann and C. R. Rao , 189–207. [MR2668352](#)
- [22] RAO, J. N. K. and WU, C. F. J. (1988). Resampling inference with complex survey data. *Journal of the American Statistical Association*, **83**, 231–241. [MR0941020](#)
- [23] RAO, J. N. K., WU, C. F. J. and YUE, K. (1992). Some recent work on resampling methods for complex surveys. *Survey Methodology*, **18**, 209–217.
- [24] REID, N. (2012). Likelihood inference in complex settings. *The Canadian Journal of Statistics*, **40**, 731–744. [MR2998859](#)
- [25] WANG, H., LI, R. and TSAI, C. L. (2007). Tuning parameter selectors for the smoothly clipped absolute deviation method. *Biometrika*, **94**, 553–568. [MR2410008](#)
- [26] WU, C. (2005). Algorithms and r codes for the pseudo empirical likelihood methods in survey sampling. *Survey Methodology*, **31**, 239–243.
- [27] WU, C. and LU, W. W. (2016). Calibration weighting methods for complex surveys. *International Statistical Review*, **84**, 79–98. [MR3491280](#)
- [28] 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](#)
- [29] ZHAO, P. and WU, C. (2019). Some theoretical and practical aspects of empirical likelihood methods for complex surveys. *International Statistical Review*, **87**, 239–256. [MR3957353](#)
- [30] ZHAO, P., HAZIZA, D. and WU, C. (2018). Empirical likelihood inference for complex surveys and the design-based oracle variable selection theory. Submitted . [MR3324425](#)
- [31] ZHAO, P., RAO, J. N. K. and WU, C. (2020). Empirical likelihood inference with public-use survey data. [arXiv:2005.12172](#).
- [32] ZHONG, B. and RAO, J. N. K. (2000). Empirical likelihood inference under stratified random sampling using auxiliary population information. *Biometrika*, **87**, 929–938. [MR1813985](#)