UNIVERSITY OF Southampton

University of Southampton Research Repository
ePrints Soton

Copyright © and Moral Rights for this thesis are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given e.g.

AUTHOR (year of submission) "Full thesis title", University of Southampton, name of the University School or Department, PhD Thesis, pagination

http://eprints.soton.ac.uk

UNIVERSITY OF SOUTHAMPTON

FACULTY OF SOCIAL AND HUMAN SCIENCES

Division of Social Statistics and Demography

Empirical likelihood confidence intervals for survey data

by

Omar De La Riva Torres

Thesis for the degree of Doctor of Philosophy

August 2014

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SOCIAL AND HUMAN SCIENCES

DIVISION OF SOCIAL STATISTICS AND DEMOGRAPHY

Doctor of Philosophy

EMPIRICAL LIKELIHOOD CONFIDENCE INTERVALS FOR SURVEY DATA

by Omar De La Riva Torres

We propose an empirical likelihood approach which can be used to construct design-based confidence intervals under unequal probability sampling without replacement with a small and large sampling fraction. It gives confidence intervals which may perform better than standard confidence intervals based on the central limit theorem and the pseudo empirical likelihood confidence intervals (Wu and Rao, 2006). It does not rely on variance estimates, design effects, joint-inclusion probabilities, resampling or linearisation. It can be applied to the Horvitz and Thompson (1952) estimator, the Hájek (1971) estimator or the regression estimator (Särndal et al., 1992). The proposed approach also offers a likelihood-based justification for design-based approaches used in sample surveys. It is less computationally intensive than bootstrap methods which can be unstable and may not have the right coverage. The proposed approach can be used to construct suitable non-parametric (design-based) confidence intervals for quantiles which do not rely on the normality, unbiasedness or linearity of the point estimator. We support our findings via simulation studies for a highly skewed population and comparing them with alternatives such as linearisation (Deville and Särndal, 1992), rescaled bootstrap (Rao and Wu, 1988; Rao et al., 1992), direct bootstrap (Antal and Tillé, 2011) and the Woodruff (1952) intervals. We apply the proposed approach to *persistent-risk-of-poverty* which is an indicator of poverty based upon the European Union Survey on Income and Living Conditions (EU-SILC, Eurostat, 2012). Confidence intervals for this indicator are estimated for sub-population domains. A description of the computational algorithms used is provided and implemented using the statistical software R (R Core Team, 2012).

Contents

List of Tables

List of Figures

Notation

$\widehat{X}_\pi$	$\widehat{X}_\pi = \sum_{i=1}^{n} \breve{x}_i$.
$\breve{x}_i$	$x_i \pi_i^{-1}$.
$\widehat{\boldsymbol{B}}$	Vector of regression coefficient estimates.
$\boldsymbol{f}(\boldsymbol{\eta})$	$\boldsymbol{f}(\boldsymbol{\eta}) = \sum_{i=1}^{n} \widehat{m}_i \boldsymbol{c}_i$.
$\widehat{\boldsymbol{C}}_\pi$	$\widehat{\boldsymbol{C}}_\pi = \sum_{i=1}^{n} \boldsymbol{c}_i \pi_i^{-1}$.
$\boldsymbol{C}^*$	$\boldsymbol{C}^* = (\boldsymbol{C}^\mathsf{T}, 0)^\mathsf{T}$.
$\boldsymbol{C}$	A known $Q \times 1$ vector.
$\boldsymbol{c}_i^*$	$\boldsymbol{c}_i^* = (\boldsymbol{c}_i^\mathsf{T}, g_i(\theta))^\mathsf{T}$.
$\boldsymbol{c}_i$	A known $Q \times 1$ vector associated with the i-th sampled unit.
$\breve{\boldsymbol{z}}_i$	$\breve{\boldsymbol{z}}_i = \boldsymbol{z}_i \pi_i^{-1}$.
$\boldsymbol{z}_i$	$\boldsymbol{z}_i = (z_{i1}, \ldots, z_{iH})^\mathsf{T}$.
$\boldsymbol{\eta}_0$	In order to compute $\boldsymbol{\eta}$ for the first iteration $\boldsymbol{\eta}_0 = \mathbf{0}$.
$\boldsymbol{\eta}$	$\boldsymbol{\eta}$ is such that (3.2.2) holds.
$\boldsymbol{\vartheta}_0$	$\boldsymbol{\vartheta}_0 = XN^{-1}$.
$\breve{\boldsymbol{f}}_i(\boldsymbol{x}_i, \boldsymbol{\vartheta}_0)$	$\breve{\boldsymbol{f}}_i(\boldsymbol{x}_i, \boldsymbol{\vartheta}_0) = \boldsymbol{f}_i(\boldsymbol{x}_i, \boldsymbol{\vartheta}_0)\pi_i^{-1}$.
$\boldsymbol{f}_i(\boldsymbol{x}_i, \boldsymbol{\vartheta}_0)$	$\boldsymbol{f}_i(\boldsymbol{x}_i, \boldsymbol{\vartheta}_0)$ is a vector function of the auxiliary variables and a known parameter $\boldsymbol{\vartheta}_0$.
$Boot$	Bootstrap approach.

$\widehat{B}_x$	Coefficient given by (3.7.11) after substituting x_i by $\pi_i - nN^{-1}$.
$\widetilde{B}_x$	Coefficient given by (3.7.11) after substituting $n^{-1}\widehat{G}_\pi(\theta)$ by $\widehat{G}_\pi(\theta)\widehat{N}_\pi^{-1}\pi_i^{-1}$ and x_i by $\pi_i - nN^{-1}$.
$\boldsymbol{B}$	Vector of regression coefficients.
$\widehat{\boldsymbol{S}}$	A positive definite covariance matrix.
$\underline{\boldsymbol{S}}_{\text{D}}$	A random sample selected with one-one design.
$\underline{\boldsymbol{S}}^*$	A random sample selected with direct bootstrap design.
$\underline{\boldsymbol{S}}$	A random sample selected with over-replacement.
$f(\cdot)$	Density function.
$\widehat{f}(\cdot)$	Estimate of a density function.
$\widehat{V}(\cdot)$	Estimator of Variance under the specified design $p(s)$.
$\widehat{\overline{Y}}$	Estimator of the population mean $\overline{Y}$.
$E(\cdot)$	Mathematical expectation.
$F^*(z)$	The inverse function of $\phi(\cdot)$.
$\widehat{F}(\cdot)$	Estimation of a cumulative distribution function.
$\widetilde{F}(\cdot)$	Smoothed cumulative distribution function.
$F(\cdot)$	Cumulative distribution function.
$\widehat{G}(\theta)$	Estimate of $G(\theta)$.
$G(\theta)$	Estimating equation.
π_{kl}	Second order inclusion probability.
$I_{(\cdot)}$	Indicator function where $I_{(\cdot)} = 1$ if the argument inside of $(\cdot)$ is true and it takes the value of 0 otherwise.
$\widehat{\boldsymbol{\Delta}}(\cdot)$	$Q \times Q$ gradient matrix.
$\boldsymbol{\lambda}$	Vector of Lagrange multipliers.

$\widehat{r}(\theta)$	Profile empirical log-likelihood ratio function.
$\ell(\cdot)$	Maximum value of the empirical log-likelihood function.
N_h	Size of stratum h, $h = 1, \ldots, H$.
N	Population size.
θ	Parameter of interest.
$\widehat{\theta}$	Estimator of θ.
θ_0	Population value of θ.
$\Phi_i(w, d)$	Distance function between w_i and the basic design weights $d_i = \pi_i$.
$\widehat{N}_\pi$	Population size estimator.
$\varrho\left(y_{(i)}, \theta\right)$	Function used in $\tilde{F}(y)$ in order to make a linear interpolation of $\widehat{F}(\theta)$.
$R(\cdot)$	Profile empirical likelihood ratio function.
$p(s)$	A given sampling design.
g_i	Calibration scale factor.
S_y^2	Population variance.
W_h	Stratum weight N_h/N.
U	A population U composed of $i = 1, \ldots, N$ units.
z_i	Linearised variable.
$V(\cdot)$	Variance under the specified design $p(s)$.
$\boldsymbol{n}$	Vector of the stratum sample size.
w_i	Sampling weights.
X	Population control total.
Y	Total of a variable of study.
$\boldsymbol{X}$	Vector of population control totals.
$\boldsymbol{x}_i$	Vector of auxiliary variables attached to the unit i.
deff	Design effect.

$\delta\{\cdot\}$	Indicator function where $\delta\{\cdot\} = 1$ if the argument inside of $\{\cdot\}$ is true and it takes the value of 0 otherwise.
d_i	Design weights $d_i = \pi_i^{-1}$.
$L(\cdot)$	Empirical likelihood function.
$\breve{g}_i(\theta)$	$\breve{g}_i(\theta) = g_i(\theta)\pi_i^{-1}$.
$\widetilde{g}_i(\theta_0)$	Residuals defined by $\widetilde{g}_i(\theta_0) = g_i(\theta_0) - \widehat{\boldsymbol{B}}^{\mathsf{T}}(Nn^{-1}\boldsymbol{z}_i, \boldsymbol{f}_i^{\mathsf{T}})^{\mathsf{T}}$.
$g_i(\theta)$	Function of θ and of the characteristics of the unit i.
iid	Independent and identically distributed variables.
λ	Lagrange multiplier.
$l(\cdot)$	Log-empirical likelihood function.
$\ell(\pi)$	$\ell(\pi) = \sum_{i=1}^{n} \log(\pi_i)$.
$r(\cdot)$	Profile log-likelihood ratio function.
m_i	The unit mass of unit i in the population.
max	Maximum of a function.
$\widehat{m}_i$	Maximum likelihood estimator of m_i.
min	Minimum of a function.
n_h	Size of the sample in stratum h.
ns	Non stratified sampling design.
n	Sample size.
$\phi(\cdot)$	$\phi_i(w,d) = \partial\Phi_i(w,d)/\partial w$.
π_i	First order inclusion probability.
ψ_i	The values ψ_i are such that the regularity conditions (3.3.4)-(3.3.8) hold for $\boldsymbol{c}_i = \boldsymbol{c}_i^*$, $\boldsymbol{C} = \boldsymbol{C}^*$.
p_i	Probability masses $p_i = Pr(Y = y_i)$.
q_i	$q_i = (1 - \pi_i)^{1/2}$.
s_h	A subset of U_h.
s	A subset of U.

$\widehat{\varsigma}$	Finite population correction proposed by Rao et al. (1962).
$\tilde{d}_i$	Normalised design weights $\sum_{i \in s} d_i = 1$.
$\widehat{Y}_\pi$	Horvitz-Thompson estimator of Y.
$\widehat{Y}_H$	Hájek estimator of Y.
x_i	Auxiliary variable attached to the i.
y_i	Values of the variable of study Y.

Acronyms

πPS	Probability proportional-to-size sampling design.
E.R.	Error rates.
EL	Empirical likelihood.
EL1	Pseudo Empirical likelihood without the additional constraint on the size measures defined in Wu and Rao (2006, p. 362).
EL2	Pseudo Empirical likelihood with the additional constraint on the size measures defined in Wu and Rao (2006, p. 362).
EU-SILC	European Statistics on Income and Living Conditions.
GREG	Generalised regression estimator.
H	Hájek estimator.
HT	Horvitz and Thompson estimator.
MU284	Swedish municipality data (Särndal et al., 1992).

PEL	Pseudo Empirical likelihood.
PPS	Probability proportional-to-size sampling design.
PSU	Primary sampling unit.
REG	Regression estimator.
REV84	Real estate values according to 1984 assessment (in millions of kronor), MU284 data.
RHC	Rao-Hartley-Cochran sampling design.
RRMSE	Relative mean square error.
SS82	The number of Social-Democratic seats in municipal council, MU284 data.
ST	Stratified sampling design.

Declaration of Authorship

I, Omar De La Riva Torres, declare that the thesis entitled *Empirical Likelihood confidence intervals for survey data* and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;

7. Either none of this work has been published before submission, or parts of this work have been published as:

   – Berger Y. G. and De La Riva Torres, O. (2011). Empirical Likelihood Ratio Confidence Intervals for Unequal Probability Sampling. In *Proceedings of International Statistical Institute 58th Congress*, Dublin: International Statistical Institute.

   – Berger, Y. G. and De La Riva Torres, O. (2012a). A unified theory of empirical likelihood ratio confidence intervals for survey data with unequal probabilities. *Proceedings of Joint Statistical Meeting*. San Diego. American Statistical Institute.

   – Berger, Y. G. and De La Riva Torres, O. (2012b). Empirical likelihood confidence intervals for complex survey designs. *Southampton Statistical Sciences Research Institute http://eprints.soton.ac.uk/337688/*.

   – Berger, Y. G. and De La Riva Torres, O. (2013). Empirical likelihood confidence intervals: An application to the EU-SILC household surveys. In *Book of Abstracts of Italian Conference on Survey Methodology 3rd Conference Edition*. Milan: libreriauniversitaria.it ed.

   – De La Riva Torres, O. and Berger Y. G. (2013). Estimation of confidence intervals for quantiles under unequal probability sampling using a new empirical likelihood approach. In *Proceedings of International Statistical Institute 59th Congress*, Hong Kong, SAR : International Statistical Institute.

Signed: ....

Date: ....

Acknowledgements

First and foremost I am grateful with my supervisor Yves G. Berger. With his support and advice since the very first day I arrived at the University of Southampton, he helped to conclude my Ph.D. studies and enjoy academic life.

I gratefully acknowledge Dankmar Böhning and Nikolaos Tzavidis for their valuable comments of a previous version of this thesis.

I would like thank the National Council for Science and Technology of Mexico (CONACYT) for the awarded scholarship during my doctoral studies; I also thank the Division of Social Statistics and Demography of University of Southampton, Maison Franco-Mexican University and the European Parliament which granted funds for conference attendance.

Thanks are also due to my parents, siblings, nieces and nephews, all of them have been particularly supportive to reach this goal.

I really appreciated the friendship and the good time I spent with Emilio López Escobar, Juan Correa Allamand, Carla Pezzulo, Rebecca Vassallo, Dimitrios Gkountanis, Luis Pinedo Caro, Nicholas Lazarou, Yahia El-Horbaty, Gregori Galofré-Vilà, Qi Zhang, Ronaldo Nazare, Andrés Luengo, Federico De Luca, Ngianga II Kandala, Sharon Holder and Dimitrios Nikolakis.

1

Introduction

Survey sampling theory has been focused on point estimation and their respective standard errors. Confidence intervals for parameters of interest have mainly been constructed assuming normality. Owen (1988) proposed the Empirical Likelihood approach as a nonparametric method of statistical inference to construct empirical likelihood ratio confidence intervals similar to the parametric likelihood ratio intervals. Unlike the bootstrap procedure where data points receive equal probabilities of selection in the sample, empirical likelihood assigns weights that maximise a constrained multinomial likelihood and empirical likelihood confidence intervals are constructed by contouring the likelihood (Hall, 1992). The empirical likelihood confidence intervals have the properties of determining the shape of confidence regions based on the observed data set and avoid inadmissible parameter values. Confidence interval coverage accuracy is at least comparable with the bootstrap confidence region, it provides balanced tail errors and is not necessarily symmetric, and it incorporates supplementary information through constraints. In addition, it is possible to apply a Bartlett correction to reduce coverage error (Hall and La Scala, 1990). The most appealing property of empirical likelihood is that the corresponding log likelihood ratio statistic is asymptotically χ^2 distributed as a true parametric likelihood.

We propose an empirical likelihood approach which can be used to construct design-based confidence intervals under unequal probability sampling without replacement. The proposed empirical likelihood confidence interval has the following advantages: it gives confidence intervals which may perform better than standard confidence intervals based on the central limit theorem. It does not rely on variance estimates, design effects or joint-inclusion probabilities. It can be applied to the Horvitz and Thompson (1952) estimator, the Hájek (1971) estimator or the regression estimator (Särndal et al., 1992). It can be also used with small and large sampling fractions. The proposed approach also offers a likelihood-based justification for design-based approaches used in sample surveys.

This thesis is integrated by the following parts:

Chapter 1. Introduction

Chapter 2. Literature review

Owen (1988) conceived and introduced the empirical likelihood method in statistical inference area. A general summary of advances and developments since its inception, when it was presented for first time in survey sampling as the *scale load* approach by Hartley and Rao (1968), is outlined. In order to give a framework to assess the proposed approach, an overview of the most frequently used methods in survey sampling for building confidence intervals is provided.

Chapter 3. Empirical likelihood ratio estimator confidence intervals for unequal probability sampling

A novel approach based on the use of empirical likelihood in survey sampling is proposed for the construction of confidence intervals for population means under unequal probability sampling design with a small and large sampling fraction. In contrast to the pseudo empirical likelihood method, our approach does not require design effect estimation and takes in account the finite population correction for a large sampling fraction.

We support our results with two simulation studies for a highly skewed population. The proposed approach does not rely on variance estimates, design-effects, resampling or linearisation.

Chapter 4. Estimation of empirical likelihood confidence intervals for quantiles

We propose a new empirical likelihood approach which can be used to construct non-parametric (design-based) confidence intervals for quantiles which do not rely on the normality of the point estimator. We show that the proposed approach gives suitable confidence intervals even when the estimator of a quantile is biased. The proposed approach also deals with large sampling fractions. The bootstrap is an alternative approach which can be used to derive non-parametric confidence intervals for quantiles. The proposed approach is less computationally intensive than the bootstrap. We compare our proposed approached with alternative approaches such as linearisation (Deville and Särndal, 1992), direct bootstrap (Antal and Tillé, 2011), rescaled bootstrap (Rao and Wu, 1988; Rao et al., 1992) and the Woodruff (1952) interval.

Chapter 5. Empirical Likelihood confidence intervals for the persistent-risk-of-poverty rate

This chapter presents an application to real data of the proposed empirical likelihood approach which can be used to construct design-based confidence intervals. The proposed approach gives confidence intervals which may have better coverage than standard confidence intervals and pseudo empirical likelihood confidence intervals (Wu and Rao, 2006), which rely on variance estimates and design-effects. The proposed approach does not rely on variance estimates, resampling or linearisation, even when the parameter of interest is not linear. We apply the proposed approach to a measure of poverty based upon the European Union Survey on Income and Living Conditions (EU-SILC, Eurostat, 2012). Confidence intervals of the poverty indicator *persistent-risk-of-poverty rate* are estimated for the overall sample and six sub-population domains (Särndal et al.,

1992) determined by three age group and sex. It also gives suitable confidence intervals when the point estimator is biased. The proposed approach is less computational intensive than rescaled (Rao and Wu, 1988; Rao et al., 1992) and direct bootstrap (Antal and Tillé, 2011) which can be unstable and may not have the right coverage.

Chapter 6. Algorithms for obtaining weights and intervals in the empirical likelihood approach

An explanation of the key computational algorithms is provided for implementation of the proposed empirical likelihood point estimators, and the construction of empirical likelihood ratio confidence intervals for the parameter of interest. The computational tasks imply the estimation of the empirical likelihood weights and empirical log-likelihood function. Note that, the crucial algorithm is the computation of the vector $\boldsymbol{\eta}$, once this quantity is obtained, the empirical likelihood weights and confidence intervals can be calculated. The computational codes associated with the algorithms are written in the statistical software R (R Core Team, 2012) and remitted to the Appendix A.

Appendix A. R functions for the empirical likelihood confidence intervals for survey data

In this appendix are described the main functions, written in statistical software R (R Core Team, 2012), used in the simulation studies of this thesis. Only the functions for the implementation of the proposed empirical likelihood approach are considered. Functions for the realization of compared approaches (standard methods based on normal distribution of data, pseudo empirical likelihood and bootstrap methods) are omitted and referred to original articles. For the sake of brevity, functions related to the estimation of empirical likelihood confidence intervals are only considered for the lower bound limit. The upper bound limit functions are similar only the direction for the seeking should be inverted. The functions including a short description are ordered alphabetically in a list.

2

Literature Review

2.1 Early Contributions in Empirical Likelihood Methodology

2.1.1 Scale load approach

Consider the population U composed of $i = 1, \ldots, N$ units with the values y_i of the variable of study Y. A subset s from U is selected with probability $p(s)$; sample data is denoted by $\{(i, y_i), i \in s\}$. The first use of EL in survey sampling was proposed by Hartley and Rao (1968) and named the scale load approach focused in the inferential aspect of point estimation. Some aspects of the sample need to be omitted to make the sample not unique and have the likelihood informative. Assuming that the variable Y is measured on a scale with finite scale points y_t^*, $t = 1, ..., T$ with N_t the number of units in U having the value y_t^*. In consequence, $N = \sum_{t \in T} N_t$ and $\overline{Y} = N^{-1} \sum_{t \in T} N_t y_t^*$ is specified by the scale loads $\mathbf{N} = (N_1, \ldots, N_T)^\mathsf{T}$. Consider n the sample size and n_t the number of units in the sample with value y_t^*, the sample data is summarised by the scale loads $\mathbf{n} = (n_1, \ldots, n_T)^\mathsf{T}$, with $n_t \geqslant 0$ and $n = \sum_{t \in T} n_t$. Under simple random sampling without replacement and negligible sample fraction n/N, the likelihood based on the reduced sample data is given by the hypergeometric distribution that depends on $\mathbf{N}$. The log-likelihood could be approximated by the multinomial log-likelihood

$$l(\mathbf{p}) = \sum_{t\in T} n_t \log(p_t) \tag{2.1.1}$$

where $\mathbf{p} = (p_1, \ldots, p_T)^\mathsf{T}$ and $p_t = N_t/N$. The maximum likelihood estimator of $\overline{Y} = \sum_{t\in T} p_t y_t^*$ is the sample mean $\bar{y} = \sum_{t\in T} \hat{p}_t y_t^*$, where $\hat{p}_t = n_t/n$. The scale load approach allows incorporation of the information of the auxiliary variable x when $\overline{X}$ is known. The scale points of x are denoted as x_j^*, $j = 1, \ldots, J$ and the scale load of (y_t^*, x_j^*) as N_{tj}. Therefore, $\overline{Y} = \sum_{t\in T}\sum_{j\in J} p_{tj} y_t^*$ and $\overline{X} = \sum_{t\in T}\sum_{j\in J} p_{tj} x_t^*$, where $p_{tj} = N_{tj}/N$ and $\sum_{t\in T}\sum_{j\in J} p_{tj} = 1$. The sample data reduces to the observed frequencies n_{tj} for the scale points (y_t^*, x_j^*) such that $\sum_{t\in T}\sum_{j\in J} n_{tj} = n$. The maximum likelihood estimator of $\overline{Y}$ is computed as $\widehat{\overline{Y}} = \sum_{t\in T}\sum_{j\in J} \hat{p}_{tj} y_t^*$ where $\hat{p}_{tj}$ maximise the log-likelihood

$$\sum_{t\in T} n_t \log(p_{tj}) \tag{2.1.2}$$

subject to the constraints

$$\sum_{t\in T}\sum_{j\in J} p_{tj} = 1 \text{ and } \sum_{t\in T}\sum_{j\in J} p_{tj} x_t^* = \overline{X}. \tag{2.1.3}$$

Hartley and Rao (1968) showed that $\widehat{\overline{Y}}$ is asymptotically equal to the customary linear regression estimator of $\overline{Y}$.

2.1.2 Empirical likelihood confidence intervals

EL was proposed by Owen (1988) considering $y_1, \ldots, y_n$ as independent and identically distributed observations from y with cumulative distribution function $F(\cdot)$. An empirical likelihood puts masses $p_i = \Pr(Y = y_i) = F(y_i) - F(y_i-)$ at the sample points, where $F(y_i-) = \lim_{v\uparrow y_i} F(v)$. The EL function is $L(\mathbf{p}) = \prod_{i\in s} p_i$. Maximising the log-likelihood

$$l(\mathbf{p}) = \log[L(\mathbf{p})] = \sum_{i\in s} \log(p_i) \tag{2.1.4}$$

under the constraints

$$p_i > 0 \text{ and } \sum_{i \in s} p_i = 1 \tag{2.1.5}$$

leads to $\hat{p}_i = 1/n$ and the maximum estimator for the mean $\mu = E(y)$ and $F(u)$ are given by $\hat{\mu} = \sum_{i \in s} \hat{p}_i y_i = \widehat{\overline{Y}}$ and $\widehat{F}(u) = \sum_{i \in s} \hat{p}_i I_{(y_i \leqslant u)} = F_n(u)$, respectively; where $I_{(y_i \leqslant u)} = 1$ if $y_i \leqslant u$ and it takes the value of 0 otherwise, and $F_n(u) = n^{-1} \sum_{i \in s} I_{(y_i \leqslant u)}$ is the empirical distribution function based on the independent and identically distributed sample.

The profile empirical likelihood ratio function $R(\mu)$, for the mean μ, is obtained by maximising

$$\frac{L(\mathbf{p})}{L(F_n(u))} = \prod_{i \in s} (np_i) \tag{2.1.6}$$

under the constraints

$$\sum_{i \in s} p_i = 1 \text{ and } \sum_{i \in s} p_i y_i = \mu. \tag{2.1.7}$$

Owen (1988) proved under mild moment conditions, $0 < V(y_i) < \infty$ and $E(|y_i|^3) < \infty$ and a suitable asymptotic framework that allows n and N simultaneously go to infinity while n/N goes to zero, $r(\mu) = -2 \log R(\mu)$ is asymptotically χ^2 with one degree of freedom. Therefore, the $1 - \alpha$ EL interval is given by $\{\mu | r(\mu) \leqslant \chi_1^2(\alpha)\}$, where $\chi_1^2(\alpha)$ is the $\alpha-$point of the χ^2 distribution with one degree of freedom.

2.2 Application in survey sampling

The first application in survey sampling using the EL approach is due to Chen and Qin (1993) assuming simple random sampling without replacement and known auxiliary information expressed as $E[w(x)] = 0$ for some known $w(\cdot)$. The parameters considered are of the form $\theta = N^{-1} \sum_{i \in N} g(y_i)$ with $g(\cdot)$ a known function and $p_i = \Pr(Y = y_i)$. The log-likelihood function is given by

$$l(\mathbf{p}) = \log[L(\mathbf{p})] = \sum_{i \in s} \log(p_i). \tag{2.2.1}$$

The maximum EL estimator of θ is obtained applying the Lagrange multipliers method and defined by $\hat{\theta} = \sum_{i \in s} \hat{p}_i g(y_i)$, where $\hat{p}_i$ maximises (2.2.1) under the constraints

$$p_i > 0 \text{ and } \sum_{i \in s} p_i = 1 \tag{2.2.2}$$

and $E[w(x_i)] = \sum_{i \in s} p_i w(x_i) = 0$. The solution obtained is

$$\hat{p}_i = \frac{1}{n[1 + \lambda w(x_i)]}$$

where the Lagrange multiplier λ is the solution to

$$\sum_{i \in s} \frac{w(x_i)}{1 + \lambda w(x_i)} = 0.$$

In the case of the selection of $g(y_i) = y_i$ and $w(x_i) = x_i - \overline{X}$, the parameter and known auxiliary information given are θ and $\sum_{i \in s} p_i x_i = \overline{X}$. Setting $g(y_i) = I_{(y_i \leqslant u)}$ for a fixed u, the maximum EL estimator of the population distribution function $\theta = F(u) = N^{-1} \sum_{i \in U} I_{(y_i \leqslant u)}$ is given by $\widehat{F}(u) = \sum_{i \in s} \hat{p}_i I_{(y_i \leqslant u)}$. The maximum EL estimator of the population quantiles can be obtained through direct inversion of $\widehat{F}(u)$ because it is a non decreasing function in the range $[0, 1]$. Chen and Qin (1993, p. 109) mentioned under their approach the estimator of $\theta = \overline{Y}$ is the post stratified estimator when $w(x_i) = x_i - \overline{X}$, x_i is a $0 - 1$ variable and $\overline{X}$ is a population proportion. Based on the similarities between the raking method (Deming and Stephan, 1940) and the post stratification method, Chen and Qin (1993) agreed that the former coincides with the EL approach.

2.3 Pseudo empirical likelihood

The estimator proposed by Chen and Qin (1993) is not possible to be generalised to any sampling design because the likelihood depends on the sampling design and a complete specification of the joint probability function is not always affordable under any without replacement sampling (Rao and Wu, 2009). The pseudo EL was suggested by Chen and Sitter (1999) claiming it was suitable for any survey design.

Let $\{y_1, \ldots, y_N\}$ be a finite population assumed to be a random sample from a super-population. The purpose is to maximise $L(\mathbf{p}) = \prod_{i \in N} p_i$, it is equivalent to maximise $\log L(\mathbf{p}) = l(\mathbf{p}) = \sum_{i \in U} \log(p_i)$. The Horvitz and Thompson (HT 1952) estimator

$$\hat{l}_{\text{HT}}(\mathbf{p}) = \sum_{i \in s} d_i \log(p_i) \tag{2.3.1}$$

of the population total of $\log(p_i)$ is used as a pseudo empirical log-likelihood with $E\left[\sum_{i \in s} d_i \log(p_i)\right] = \sum_{i \in U} \log p_i$, where $d_i = \pi_i^{-1}$ are the design weights and π_i are the inclusion probabilities. The maximum pseudo EL estimator of $\overline{Y}$, which is determined by the maximisation of (2.3.1) subject to the constraints in (2.2.2) and using the Lagrange multiplier method, corresponds to the Hájek (1964) estimator of $\overline{Y}$ as $\overline{Y}_{\text{H}} = \sum_{i \in s} \tilde{d}_i y_i$, where $\tilde{d}_i = d_i / \sum_{i \in s} d_i$. Though, the Hájek estimator is less efficient than the Horvitz-Thompson estimator $\widehat{\overline{Y}} = N^{-1} \sum_{i \in s} d_i y_i$ under probability proportional to size sampling (PPS) without replacement with π_i proportional to the size x_i when y_i is proportional to x_i. In the case of known population mean $\overline{X}$ of the auxiliary variable included in the constraint

$$\sum_{i \in s} p_i x_i = \overline{X}.$$

For a comparison between pseudo EL and calibration, consider a known population mean $\overline{\boldsymbol{X}}$ of a vector of auxiliary variables $\boldsymbol{x}$. The maximum pseudo EL estimator of $\overline{Y}$ is $\widehat{\overline{Y}}_{\text{PEL}} = \sum_{i \in s} \hat{p}_i y_i$ where $\hat{p}_i$ maximise $l(\mathbf{p})$ subject to $p_i > 0$, $\sum_{i \in s} p_i = 1$ and

$$\sum_{i \in s} p_i \boldsymbol{x}_i = \overline{\boldsymbol{X}}. \tag{2.3.2}$$

The Lagrange multiplier method leads to the solution $\hat{p}_i = \tilde{d}_i(s)/(1 + \boldsymbol{\lambda}^{\mathsf{T}} \mathbf{u}_i)$ where $\mathbf{u}_i = \boldsymbol{x}_i - \overline{\boldsymbol{X}}$ and $\boldsymbol{\lambda}$ is the solution to

$$g(\boldsymbol{\lambda}) = \sum_{i \in s} \frac{\tilde{d}_i(s) \mathbf{u}_i}{1 + \boldsymbol{\lambda}^{\mathsf{T}} \mathbf{u}_i} = \mathbf{0}. \tag{2.3.3}$$

A calibration estimator of $\overline{Y}$ is defined by $\overline{Y}_{\text{C}} = N^{-1}\sum_{i\in s} w_i y_i$ where the calibrated weights w_i minimise the distance between $\mathbf{w} = (w_1, \ldots, w_n)^\mathsf{T}$ and the basic design weights $\mathbf{d} = (d_1, \ldots, d_n)^\mathsf{T}$ subject to $\sum_{i\in s} \mathbf{w}_i \boldsymbol{x}_i = \boldsymbol{X}$.

The chi square distance function $\Phi(\mathbf{w}, \mathbf{d}) = \sum_{i\in s} (w_i - d_i)^2/d_i q_i$ with q_i predefined, produces a generalised regression (GREG) estimator of $\overline{Y}$ (Särndal et al., 1992). However, the calibrated weights can be negative and the use of other distance functions that forces the weights to be positive involve other disadvantages such as there is no a guarantee of finding a solution that minimises the distance functions, or some extreme large weights w_i may arise (Deville and Särndal, 1992). The pseudo EL likelihood approach to calibration estimation has two appealing features: (i) the weights are positive and normalised is attractive for the maximum pseudo EL estimator of $F_{\text{N}} = N^{-1}\sum^{i\in N} I_{(y_i \leq y)}$ computed as $\widehat{F}_{\text{PEL}}(t) = \sum_{i\in s} \hat{p}_i I_{(y_i \leqslant t)}$, because it is a genuine distribution function and by the inversion of $\widehat{F}_{\text{PEL}}(t)$ quantile estimates can be obtained; (ii) for the most laborious computational task, the finding of the Lagrange multipliers λ, a Newton-Raphson algorithm can be used (Chen et al., 2002).

Under the following regularity conditions

C1. $\max\limits_{i\in s} |u_i| = o_p(n^{1/2})$

C2. $\dfrac{\sum_{i\in s} d_i(s) u_i}{\sum_{i\in s} d_i(s) u_i^2} = O_p(n^{-1/2})$

the solution is given by $\boldsymbol{\lambda} = \left(\sum_{i\in s} d_i \mathbf{u}_i \mathbf{u}_i^\mathsf{T}\right)^{-1} \sum_{i\in s} d_i \mathbf{u}_i + o_p(n^{-1/2})$ and $\hat{p}_i \approx \tilde{d}_i(s)(1 - \boldsymbol{\lambda}^\mathsf{T}\mathbf{u}_i)$, which implies that

$$\widehat{\overline{Y}}_{\text{PEL}} = \widehat{\overline{Y}}_{\text{H}} + \widehat{\boldsymbol{B}}^\mathsf{T}(\overline{\boldsymbol{X}} - \widehat{\overline{\boldsymbol{X}}}_{\text{H}}) + o_p(n^{-1/2}) \tag{2.3.4}$$

where $u_i = x_i - \overline{X}$, $\mathbf{u}_i = \boldsymbol{x}_i - \widehat{\overline{\boldsymbol{X}}}_{\text{H}}$, $\widehat{\boldsymbol{B}} = \left(\sum_{i\in s} d_i \mathbf{u}_i \mathbf{u}_i^\mathsf{T}\right)^{-1} \sum_{i\in s} d_i \mathbf{u}_i y_i$, $\widehat{\overline{\boldsymbol{X}}}_{\text{H}} = \sum_{i\in s} \tilde{d}_i(s) \boldsymbol{x}_i$, and $\widehat{\overline{Y}}_{\text{H}} + \widehat{\boldsymbol{B}}^\mathsf{T}(\overline{\boldsymbol{X}} - \widehat{\overline{\boldsymbol{X}}}_{\text{H}})$ is a GREG estimator of $\overline{Y}$. From (2.3.4) it follows that any consistent variance estimator for the GREG will continue consistent for $\widehat{\overline{Y}}_{\text{PEL}}$. Although the use of a variance estimator for the GREG is asymptotically valid, it is better to apply methods of variance estimation such as linearisation, jackknife or bootstrap directly to $\widehat{\overline{Y}}_{\text{PEL}}$, recalculating $\hat{p}_i$, and not to the GREG which approximate it

(Chen and Sitter, 1999). The same statement holds in the estimation of the variance of $\widehat{F}_{\text{PEL}}(t)$ replacing y_i by $I_{(y_i \leqslant t)}$ in (2.3.4).

The pseudo EL function $\hat{l}_{\text{HT}}(\mathbf{p})$ requires only the first order inclusion probabilities and does not capture the design effects under general unequal probability sampling without replacement. As an alternative to manage these limitations Wu and Rao (2006) proposed the next pseudo empirical log likelihood function under non stratified (ns) sampling design

$$l_{ns}(\mathbf{p}) = n \sum_{i \in s} \tilde{d}_i(s) \log(p_i), \tag{2.3.5}$$

where $\tilde{d}(s) = d_i / \sum_{i \in s} d_i$. The pseudo EL function is directly related to the Kullback-Leibler distance (DiCiccio and Romano, 1989 cited in Rao and Wu, 2009) between $\mathbf{p} = (p_1, \ldots, p_n)^\mathsf{T}$ and $\tilde{\mathbf{d}} = [\tilde{d}_1(s), \ldots, \tilde{d}_n(s)]^\mathsf{T}$ in the form of $D(\mathbf{d}(s), \mathbf{p}) = \sum_{i \in s} \tilde{d}_i(s) \log(\tilde{d}_i(s)/p_i) = \sum_{i \in s} \tilde{d}_i(s) \log(\tilde{d}_i(s)) - l_{ns}(\mathbf{p})/n$, minimising this distance with respect to p_i under a given set of constraints is equivalent to maximising the pseudo EL function with respect to the same set of constraints.

The pseudo EL function uses the normalised weights $\tilde{d}_i(s)$ instead the d_i applied in $l_{\text{HT}}(\mathbf{p})$. Though maximising (2.3.5) subject to a set of constraints on the p_i is equivalent to maximizing $l_{\text{HT}}(\mathbf{p})$ subject to the same set of constraints and the maximum pseudo EL estimators obtained are the same but $l_{ns}(\mathbf{p})$ needs an adjustment in the design effect in the construction of confidence intervals (see Section 2.5.1). Without auxiliary information, maximizing $l_{ns}(\mathbf{p})$ subject to $p_i > 0$ and $\sum_{i \in s} p_i = 1$ the solution is $\hat{p}_i = \tilde{d}_i(s)$. The maximum pseudo EL estimator of $\overline{Y}$ is defined as $\widehat{\overline{Y}}_{\text{PEL}} = \sum_{i \in s} \hat{p}_i y_i$ which coincides with the Hájek estimator $\widehat{\overline{Y}}_{\text{H}} = \sum_{i \in s} \tilde{d}_i(s) y_i$ and the maximum pseudo EL estimator of the distribution function $F_{\text{N}}(t)$ is given by $\widehat{F}_{\text{H}}(t) = \sum_{i \in s} d_i I_{(y_i \leqslant t)} / \sum_{i \in s} d_i$. By imposing the constraint

$$\sum_{i \in s} p_i \pi_i = \frac{n}{N} \tag{2.3.6}$$

an improved estimator can be obtained compared with the Hájek estimator. $\sum_{i \in s} p_i z_i = \overline{Z}$ is the same as (2.3.6), where z_i is the size variable and $\overline{Z}$ is the known population mean. Chen and Sitter (1999) showed the maximum pseudo EL estimator is asymptotically equivalent to a regression type estimator with asymptotic variance equivalent to the asymptotic variance of the Hájek estimator depending on the residuals $r_i = y_i - \overline{Y} - B(z_i - \overline{z})$, where $B = \sum_{i=1}^{N}(z_i - \overline{Z})y_i / \sum_{i=1}^{N}(z_i - \overline{Z})^2$. The Hájek estimator of $F_{\text{N}}(t)$ at a fixed t has the advantage of $I_{(y_i \leqslant t)}$ is weakly correlated with the size variable z_i and $\widehat{F}_{\text{H}}(t)$ is itself a distribution function. Direct inversion of $F_{\text{N}}(t)$ can be used for the estimation of population quantiles.

2.4 Empirical Likelihood Methods in Stratified Sampling

Stratification in survey sampling is a common practice because of suitability in different situations: administrative reasons (when the territory is divided in geographic districts), the strata could be very different between them and a potential gain in efficiency in comparison with other non stratified designs. In stratified sampling design the population is divided into H finite number of non overlapped strata denoted $U_1, \ldots, U_H$, of size $N_1, \ldots, N_H$, $W_h = N_h/N$ the stratum weight with $\sum_{h=1}^{H} W_h = 1$ and $\sum_{h=1}^{H} N_h = N$; in each stratum h a probability sample $s_h \in S_h$ is selected, where S_h is the set of all possible samples of size n_h. Zhong and Rao (1996, 2000) proposed the application of EL under stratified survey sampling, considering $\{(y_{hi}, x_{hi}), i \in s_h, h = 1, \ldots, H\}$ as the sample data from the stratified sampling design. The log-likelihood

$$l(\underline{\mathbf{p}}) = l(\mathbf{p}_1, \ldots, \mathbf{p}_H) = \sum_{h=1}^{H} \sum_{i \in s_h} \log p_{hi} \qquad (2.4.1)$$

assuming negligible sampling fraction within strata and $\mathbf{p}_h = (p_{h_1}, \ldots, p_{h_{n_h}})^{\mathsf{T}}$ and p_{h_i} is the probability mass assigned to y_{hi}. Assume that population mean $\overline{X}$ of the auxiliary variable is known but $\overline{X}_h$ is unknown. If the latter information is available then the approach of Chen and Qin (1993) could be used in each stratum. The maximum EL estimator of the population mean $\overline{Y}$ is given by $\widehat{\overline{Y}} = \sum_{h=1}^{H} W_h \sum_{i \in s_h} \hat{p}_{hi} y_{hi}$, where $\hat{p}_{hi}$

are obtained by maximising (2.4.1) subject to the constraints

$$p_{hi} > 0,\ \sum_{i\in s_h} p_{hi} = 1 \text{ and } \sum_{h=1}^{H} W_h \sum_{i\in s_h} p_{hi}x_{hi} = \overline{X}. \tag{2.4.2}$$

For the case of the pseudo EL approach, Chen and Sitter (1999) proposed a method for stratified random sampling with known stratum weights W_h. Let $l(\underset{\bullet}{\mathbf{p}}) = \sum_{h=1}^{H}\sum_{i\in U_h} \log p_{h_i}$ be the population log-likelihood, the HT estimator of $l(\underset{\bullet}{\mathbf{p}})$ under stratified sampling is given by

$$\widehat{l}_{\text{HT}}(\underset{\bullet}{\mathbf{p}}) = \sum_{h=1}^{H}\sum_{i\in s_h} d_{hi}\log p_{hi} \tag{2.4.3}$$

where d_{hi} are the design weights such that $E[\sum_{h=1}^{H}\sum_{i\in s_h} d_{hi}\log p_{h_i}] = \sum_{h=1}^{H}\sum_{i\in U_h}\log p_{hi}$, in the case of stratified random sampling $d_{hj} = N_h/n_h$. The maximum EL of $\widehat{l}_{\text{HT}}(\underset{\bullet}{\mathbf{p}})$ is obtained by maximising (2.4.3) subject to the constraints (2.4.2). The maximum EL estimator of $\overline{Y}$ is given by $\widehat{\overline{Y}} = \sum_{h=1}^{H} W_h \sum_{i\in s_h}\widehat{p}_{hi}y_{hi}$. Wu and Rao (2006) used a different pseudo EL function

$$\widehat{l}_{\text{HT}}(\mathbf{p}^{\circ}) = \sum_{h=1}^{H} W_h \sum_{i\in s_h} d^{*}_{hi}\log p_{hi}/\text{deff} \tag{2.4.4}$$

subject to the same constraints (2.4.2) and where $d^{*}_{hi} = d_{hi}/\sum_{i\in s_h} d_{hi}$ the normalised design weights for stratum h and the (deff) design effect (in Section 2.5.1 its computation is explained). The value of the deff is not required for the point estimation but it is needed for the construction of confidence intervals.

2.5 Empirical Likelihood Ratio Confidence Intervals

The EL approach provides non-parametric confidence intervals for parameter of interest. The attractive properties of the EL are that its shape and orientation are determined by the data through profiling the EL ratio function. A $100(1-\alpha)\%$ empirical likelihood confidence interval for the population parameter μ_0 is given by the values μ for which

$\{\mu : r(\mu) \leqslant \chi^2_{1-\alpha,1}\}$, where $\chi^2_{1-\alpha,1}$ is the $(1-\alpha)$th quantile of the chi-square distribution with one degree of freedom. The confidence interval is built taking all the values μ such that $-2\log R(\mu) \leqslant \chi^2_{1,1-\alpha}$ holds. The probability that the population value μ_0 is in the interval approaches the nominal value $1-\alpha$ as $n \to \infty$ (Owen, 2001, p. 17). Then, the coverage error $\Pr(-2\log R(\mu_0) \leqslant \chi^2_{1,1-\alpha}) - (1-\alpha) \to 0$ as $n \to \infty$. The EL confidence intervals are asymptotic confidence intervals, and do not require estimation of standard errors or design effects.

Chen et al. (2003) estimated confidence intervals for the mean of populations containing many zero values. In this section the notation used is the conventional: let $y_1, \ldots, y_n$ be a sample of independent and identically distributed (iid) random variables with common distribution function $F(\cdot)$ and p_i the probability of observing y_i. The empirical log-likelihood for the mean $\overline{Y}$ is defined as

$$l(\mathbf{p}(\overline{Y})) = \sum_{i\in s} \log p_i,\ p_i \geqslant 0,\ \sum_{i\in s} p_i = 1,\ \sum_{i\in s} p_i y_i = \overline{Y}$$

After profiling p_i's using the Lagrange multiplier λ, the log-likelihood obtained is

$$l(\mathbf{p}(\overline{Y})) = -\sum_{i\in s} \log[1+\lambda(y_i-\overline{Y})] - n\log n \tag{2.5.1}$$

where λ solves the equation

$$\sum_{i\in s} \frac{y_i-\overline{Y}}{1+\lambda(y_i-\overline{Y})} = 0.$$

The profile EL ratio function is

$$r(\overline{Y}) = 2\left\{\max l(\mathbf{p}(\overline{Y})) - \max l(\mathbf{p})\right\} = 2\sum_{i\in s} \log[1+\lambda(y_i-\overline{Y})]. \tag{2.5.2}$$

Owen (1990) showed that for the case of $y_1, \ldots, y_n$ iid variables, where $\overline{Y}$ is the true mean parameter, the third moment of y_i exists and as $n \to \infty$, $r(\overline{Y})$ converges to the $\chi^2_{1,1-\alpha}$. Therefore, the confidence interval for $\overline{Y}$ is given by $\{\overline{Y} : r(\overline{Y}) \leqslant \chi^2_{1,1-\alpha}\}$.

In the pseudo EL approach the confidence intervals have to be adjusted for the design effect (Wu and Rao, 2006) whose definition depends on the probability sampling design and the auxiliary information used. For the non stratified sampling design, the pseudo EL ratio function for $\overline{Y}$ without considering any auxiliary information is

$$r_{ns} = -2\{l_{ns}(\hat{\mathbf{p}}(\theta)) - l_{ns}(\hat{\mathbf{p}})\}, \tag{2.5.3}$$

where $\hat{p}_i = \tilde{d}_i(s)$ maximise $l_{ns}(\mathbf{p})$ given by (2.3.5) subject to $p_i > 0$ and $\sum_{i \in s} p_i = 1$, and $\hat{p}_i(\theta)$ are values of p_i obtained by maximising $l_{ns}(\mathbf{p})$ subject to the previous restriction and the additional one $\sum_{i \in s} p_i y_i = \theta$ for a fixed θ. The deff is associated with the estimator $\widehat{\overline{Y}}_{\text{H}}$ and defined by

$$\text{deff}_{\text{H}} = \frac{V_p(\widehat{\overline{Y}}_{\text{H}})}{S_y^2/n}, \tag{2.5.4}$$

where S_y^2 is the population variance, S_y^2/n is the variance of $\widehat{\overline{Y}}_{\text{H}}$ assuming simple random sampling and ignoring the finite population correction factor n/N, and $V_p(\cdot)$ denotes the variance under the specified design $p(s)$. Under these three regularity conditions:

C1. The sampling design $p(s)$ and the study variable y satisfy $\max_{i \in s} |y_i| = o_p(n^{1/2})$.

C2. The sampling design $p(s)$ satisfies $N^{-1} \sum_{i \in s} d_i - 1 = O_p(n^{-1/2})$.

C3. The HT estimator $\hat{\theta}_{\text{HT}} = N^{-1} \sum_{i \in s} d_i y_i$ of $\theta_0 = \overline{Y}$ is asymptotically normally distributed.

The pseudo EL ratio function $r_{ns}(\theta)$ converges in distribution to a scaled χ_1^2 random variable when $\theta = \overline{Y}$. Condition C1 states that for random variables with finite variance the largest value in a sample of size n can not grow to infinity as fast as $n^{1/2}$ (Owen, 2001). Condition C2 says that $\widehat{N} = \sum_{i \in s} d_i$ is a $\sqrt{n}$ consistent estimator of N and Condition C3 is the central limit theorem for the Horvitz-Thompson estimator (Wu and Rao, 2006). They also proved that the scale factor is equal to the sampling design deff_{H}. Therefore, the adjusted pseudo EL function is $r_{ns}^{[a]}(\theta) = r_{ns}(\theta)/\text{deff}_{\text{H}}$ is asymptotically distributed as χ_1^2 when $\theta = \overline{Y}$.

A $(1 - \alpha)\%$ confidence interval for $\overline{Y}$ is defined for the maximum and minimum of the set

$$\{\theta | r_{ns}^{[a]}(\theta) \leqslant \chi_1^2(\alpha)\}, \tag{2.5.5}$$

where $\chi_1^2(\alpha)$ is the upper α quantile of the χ_1^2 distribution. The determination of the confidence interval requires a profiling analysis. For the cases of non stratified sampling designs and the use of auxiliary information $\boldsymbol{x}$ with vector of known population means $\overline{\boldsymbol{X}}$, the pseudo EL ratio function is defined similarly as in (2.5.3) with the constraint $\sum_{i\in s} p_i \mathbf{x}_i = \overline{\mathbf{X}}$ included in the determination of $\hat{p}_i$ and $\hat{p}_i(\theta)$. Under these conditions the design effect associated with the estimator $\widehat{\overline{Y}}_{\text{GREG}}$ is defined as

$$\text{deff}_{\text{GREG}} = \frac{V_p(\widehat{\overline{Y}}_{\text{GREG}})}{(S_r^2/n)} \tag{2.5.6}$$

where $V_p(\widehat{\overline{Y}}_{\text{GREG}}) = V_p\left[\sum_{i\in s} \tilde{d}_i(s) r_i\right]$, $r_i = y_i - \overline{Y} - \boldsymbol{B}^{\mathsf{T}}\mathbf{u}_i$,

$$\boldsymbol{B} = \left(\sum_{i\in U} d_i \mathbf{u}_i {\mathbf{u}^{\mathsf{T}}}_i\right)^{-1} \sum_{i\in U} d_i \mathbf{u}_i y_i,$$

$\mathbf{u}_i = \boldsymbol{x}_i - \overline{\boldsymbol{X}}$, $S_r^2 = (N-1)^{-1}\sum_{i\in U} r_i^2$. Under the assumption that conditions C1 and C3 apply to x and conditions C1, C2 and C3 hold, the adjusted PEL ratio statistic is $r_{ns}^a(\theta) = r_{ns}(\theta)/\text{deff}_{\text{GREG}}$ is asymptotically distributed as χ_1^2 when $\theta = \overline{Y}$.

EL intervals on the population for stratified sampling design was studied by Zhong and Rao (2000). The log-likelihood ratio function is given by

$$r_{\text{ST}}(\theta) = -2\left[l_{\text{ST}}(\hat{\underline{\mathbf{p}}}(\overline{Y})) - l_{\text{ST}}(\hat{\underline{\mathbf{p}}})\right] \tag{2.5.7}$$

where $\underline{\mathbf{p}}(\overline{Y})$ is the maximum likelihood estimator of $\underline{\mathbf{p}}$ subject to the constraints (2.4.2) with the additional constraint

$$\sum_{h=1}^{H} W_h \sum_{i\in s_h} p_{hi} y_{hi} = \theta. \tag{2.5.8}$$

The empirical log-likelihood ratio function obtained was adjusted to take into account the sampling fractions within strata, the adjusted function is asymptotically χ_1^2. For proportional sample allocation sampling to the strata the adjustment factor reduces to $1 - n/N$.

The pseudo EL ratio function of $\overline{Y}$ under stratified sampling with known population mean $\overline{\boldsymbol{X}} = \sum_{h=1}^{H} W_h \overline{\boldsymbol{X}}_h$ and unknown strata means $\overline{\boldsymbol{X}}_h$ is defined by

$$r_{st}(\theta) = -2\left[l_{st}(\hat{\mathbf{p}}_1(\theta), \ldots, \hat{\mathbf{p}}_H(\theta)) - l_{st}(\hat{\mathbf{p}}_1, \ldots, \hat{\mathbf{p}}_H)\right] \tag{2.5.9}$$

where $\hat{p}_{hi}$ maximise $l_{st}(\hat{\mathbf{p}}_1, \ldots, \hat{\mathbf{p}}_H)$ defined by (2.4.1) subject to the constraints (2.4.2) and $\hat{p}_{hi}(\theta)$ maximise $l_{st}(\mathbf{p}_1, \ldots, \mathbf{p}_H)$ subject to (2.4.2) with the additional constraint (2.5.7) for a fixed θ. Under a set of regularity conditions on the sampling design and variables considered within each stratum, the adjusted pseudo EL ratio statistic

$$r_{\text{ST}}^{[a]}(\theta) = r_{\text{ST}}(\theta)/\text{deff}_{\text{GREG(ST)}}$$

is asymptotically distributed as χ_1^2 when $\theta = \overline{Y}$ when θ. Where the design effect is defined as

$$\text{deff}_{\text{GREG(ST)}} = \frac{\sum_{h=1}^{H} W_h^2 V_p \left(\sum_{i \in s_h} \tilde{d}_{hi}(s_h) r_{hi}\right)}{\left(\frac{S_r^2}{n}\right)} \tag{2.5.10}$$

where $r_{hi} = (y_{hi} - \overline{Y}) - \boldsymbol{B}^{*\mathsf{T}}(\boldsymbol{x}_{hi}^* - \overline{\boldsymbol{X}}^*)$, where $\boldsymbol{B}^*$ is the population vector of regression coefficients defined as $\boldsymbol{B}$ for $\text{deff}_{\text{GREG}}$ defined in (2.5.7) using $S_r^2 = (N-1)^{-1} \sum_{h=1}^{H} \sum_{i \in U_h} r_{hi}^2$ and using $\overline{\boldsymbol{X}}^*$ to denote the augmented $\overline{\boldsymbol{X}}$ to include $W_1, \ldots, W_{H-1}$ as its first $H-1$ components and $\boldsymbol{x}_i^*$ to denote the augmented $\boldsymbol{x}_{hi}$ to include the first $H-1$ stratum indicator variables, the set of constraints can be redefined as

$$\sum_{h=1}^{H} W_h \sum_{i \in s_h} p_{hi} = 1 \text{ and } \sum_{h=1}^{H} W_h \sum_{i \in s_h} p_{hi} {\boldsymbol{x}^*}_{hi} = \overline{\boldsymbol{X}}^*. \tag{2.5.11}$$

2.5.1 Algorithm for the Construction of Confidence Intervals

The computational algorithms for seeking the upper and lower bound of the EL confidence interval employ the same profile analysis for non stratified and stratified sampling design. The construction of EL or pseudo EL ratio confidence intervals for $\theta_0 = \overline{Y}$ requires two phases: in the first one, calculate the profile EL ratio function $r(\theta)$ for a given θ and in the second one, find the lower and upper bounds for $\{\theta | r(\theta) \leqslant \chi_{1,1-\alpha}^2\}$.

In the first phase the constraint $\sum_{i\in s} p_i y_i = \theta$ for a given θ as and additional component for the set of constraints $\sum_{i\in s} p_i \boldsymbol{x}_i = \overline{\boldsymbol{X}}$. The lower and upper bound can be found via a bisection search method between $y_{(1)} = \min_{i\in s} y_i$ and $y_{(n)} = \max_{i\in s} y_i$, and the EL ratio function $r(\theta)$ is monotonically decreasing for $\theta \in (y_{(1)}, \widehat{\overline{Y}}_{\text{EL}})$ and monotonically increasing for $\theta \in (\widehat{\overline{Y}}_{\text{EL}}, y_{(n)})$. Wu (2004) showed that $r(\theta)$ is a concave function of θ and then it is maximized when $\theta = \widehat{\overline{Y}}_{\text{EL}}$. The same procedure is applied for the construction of EL for stratified sampling and pseudo EL ratio confidence intervals.

2.6 Other approaches in Empirical Likelihood

This section describes some extensions and applications for EL method in survey sampling that deserve attention and it serves to suggest new ideas and improvements.

2.6.1 Calibration estimator under unequal probability sampling

Calibration estimation using EL in survey sampling proposed by Chen and Qin (1993) was extended by Kim (2009) for unequal probability sampling under Poisson sampling. From a vector $\{y_1, \ldots, y_N\}$ of realized values of a finite population, a sample $\{y_1, \ldots, y_n\}$ is selected as the result of N Bernoulli trials where $\pi_i = \pi(Y_i)$ are the probabilities of selecting the unit i. Kim (2009) used y to denote the sample value and Y to denote the population values. The sample distribution function under Poisson sampling is defined as

$$\text{Pr}(y \leqslant x) = \text{Pr}\left[Y \leqslant x | U \leqslant \pi(Y)\right] = \frac{\sum_{(i;y_i \leqslant x)} \pi_i w_i}{\sum_{j\in s} \pi_j w_j} \tag{2.6.1}$$

where w_i is the amount of point mass that unit y_i represents in the population and can be written as $w_i = F_0(y_i) - F_0(y_i-)$, $F_0(\cdot)$ is the true distribution function and U has a uniform distribution in $(0, 1)$. The EL under Poisson sampling is determined as

$$L(w) = \prod_{i\in s} \left(\frac{\pi_i w_i}{\sum_{j\in s} \pi_j w_j} \right) \tag{2.6.2}$$

subject to $\sum_{i \in s} w_i = 1$ and $w_i \geqslant 0$. The maximum likelihood estimator of w_i is

$$\widehat{w}_i = \frac{\pi_i^{-1}}{\sum_{j \in s} \pi_j^{-1}}. \tag{2.6.3}$$

The maximum likelihood estimator for the population mean $\overline{Y}$ is defined as $\widehat{\overline{Y}} = \sum_{i \in s} \widehat{w}_i y_i$, which is equal to the Hájek estimator of the population mean. The log-likelihood $l(w) = \log\{L(w)\} = \sum_{i \in s} \log(w_i)$. Maximising $l(w)$ subject to $w_i \geqslant 0$, $\sum_{i \in s} w_i = 1$ and $\sum_{i \in s} w_i x_i = \overline{X}$ where $\overline{X}$ is the known population mean of the auxiliary variable x_i in the sample, the EL calibration estimator for $\overline{Y}$ is obtained as $\widehat{\overline{Y}}_c = \sum_{i \in s} \widehat{w}_i y_i$. This estimator has to be approximated using the Lagrange multiplier method. The solution is given by

$$\widehat{w}_i = \frac{1}{n\left[\lambda_1 \tilde{\pi}_i + \lambda_2 (x_i - \overline{X})\right]} \tag{2.6.4}$$

where $\tilde{\pi}_i = (\widehat{N}/n)\pi_i$ with $\widehat{N} = \sum_{i \in s} \pi_i^{-1}$ and λ_1 and λ_2 are the solutions to

$$\sum_{i \in s} \frac{(x_i - \overline{X})}{\lambda_1 \tilde{\pi}_i + \lambda_2 (x_i - \overline{X})} = 0. \tag{2.6.5}$$

The estimator $\sum_{i \in s} \widehat{w}_i y_i$ under a set of assumptions is asymptotically equal to the GREG estimator of $\overline{Y}$. However, unlike the GREG estimator, the associated weights $\widehat{w}_i$ in the maximum likelihood estimator $\sum_{i \in s} \widehat{w}_i y_i$ are always positive.

This approach can be extended to stratified sampling with unequal probabilities of selection in each stratum. In this case the population of size N is partitioned in H strata of sizes $N_1, \ldots, N_H$. For the stratum h with probability π_{hi}, y_{hi} and x_{hi} are observed, for $i = 1, \ldots, n_h$ and only the population mean $\overline{X}$ of x_{hi} is known. The proportion that unit y_{hi} represents in the population in the stratum h is represented by w_{hi}. The maximum likelihood estimator of $\overline{Y}$ is obtained by maximizing

$$L(\underline{\mathbf{w}}) = L(\mathbf{w}_1, \ldots, \mathbf{w}_H) = \prod_{h=1}^{H} \prod_{i \in s_h} \left(\frac{\pi_{hi} w_{hi}}{\sum_{i \in s_h} \pi_{hj} w_{hj}} \right) \tag{2.6.6}$$

subject to

$$\sum_{i \in s_h} w_{hi} = 1, \ \sum_{h=1}^{H} W_h \sum_{i \in s_h} w_{hi} x_{hi} = \overline{X}, \ h = 1, \ldots, H \tag{2.6.7}$$

and $w_{hi} \geqslant 0$, where $W_h = N_h/N$. The Lagrange multiplier method provides the solution

$$\widehat{w}_{hi} = \frac{1}{n_h \left[\lambda_h \tilde{\pi}_{hi} + \lambda_{H+1} m_h (x_{hi} - \tilde{x}_h)\right]} \tag{2.6.8}$$

where $\tilde{\pi}_{hi} = (\widehat{N}_h/n_h)\pi_{hi}$, $\widehat{N}_h = \sum_{i \in s_h} \pi_{hi}^{-1}$, $\tilde{x}_h = \sum_{i \in s_h} \widehat{w}_{hi} x_{hi}$ and $\lambda_h, h = 1, \ldots, H, H+1$, are the solutions to

$$\sum_{i \in s_h} \frac{(x_{hi} - \tilde{x}_h)}{\lambda_h \tilde{\pi}_{hi} + \lambda_{H+1} m_h (x_{hi} - \tilde{x}_h)} = 0. \tag{2.6.9}$$

Kim (2009) has shown that the maximum likelihood estimator of population mean $\overline{Y}$, for a sequence of stratified populations and samples with H fixed, is asymptotically equivalent to an optimal linear regression estimator proposed by Rao (1994). In the case when stratified random sampling is used, it is equivalent to the estimator of Zhong and Rao (2000).

2.6.2 Jackknife empirical likelihood

Jing et al. (2009) pointed out that Owen (1988)'s EL has a straightforward application when the maximisation of the non parametric likelihood involve linear constraints and the problem reduces to solve a fixed number of simultaneous equations, independent of the sample size n. Though, in applications including nonlinear statistics, the maximisation problem requires solving a set of simultaneous equations dependent on the sample size n. A $U-$statistic was used to illustrate the difficulties of applying Owen (1988)'s EL for non linear statistics. To handle the nonlinear constraint problem they proposed the combination of nonparametric approaches: jackknife and EL. The basic idea is to transform the parameter of interest into a sample mean based on jackknife pseudo values. Owen (1988)'s EL can be applied to the mean of jackknife pseudo values if they are asymptotically independent. The jackknife EL method assumes only independent random variables $y_1, \ldots, y_n$. A consistent estimator of the parameter θ is defined by $T_n = T(y_1, \ldots, y_n)$ and jackknife pseudo values $\widehat{V}_i = nT_n - (n-1)T_{n-1}^{(-i)}$, where $T_{n-1}^{(-i)} = T(y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n)$ is the statistic T_{n-1} computed with the original sample by deleting the ith observation. The jackknife estimator of θ is the average of the pseudo values $\widehat{T}_{n,jackknife} = n^{-1}\sum_{i\in s}\widehat{V}_i$.

Shi (1984, cited in Jing et al. 2009) showed that the pseudo values are asymptotically independent under mild conditions. Therefore, the jackknife estimator $\widehat{T}_{n,jack}$ for θ is a sample average of approximately independent random variables $\widehat{V}_i$ and the EL approach is suitable to be applied to the pseudo-values $\widehat{V}_i$. By defining the usual conditions in EL as $\mathbf{p} = (p_1, \ldots, p_n)^\mathsf{T}$ such that $\sum_{i\in s} p_i = 1$, $p_i \geqslant 0$ and $G_n(x) = \sum_{i\in s} p_i I_{(\widehat{V}_i \leqslant x)}$. Consider the mean functional $\theta(G_n) = \sum_{i\in s} p_i \widehat{V}_i$ and $\theta_{\mathbf{p}} = \sum_{i\in s} p_i E[\widehat{V}_i]$. Thus, the EL evaluated at θ is given by $L(\theta) = \max\left\{\prod_{i\in s} p_i | \sum_{i\in s} p_i = 1, \theta(G_n) = \theta_{\mathbf{p}}\right\}$ which attains the maximum n^{-n} at $\hat{p}_i = 1/n$. The jackknife EL ratio at θ is

$$R(\theta) = \frac{L(\theta)}{n^{-n}} = \max\left\{\prod_{i\in s} p_i \Big| \sum_{i\in s} p_i = 1, \theta(G_n) = \theta_{\mathbf{p}}\right\}. \tag{2.6.10}$$

The EL method uses Lagrange multipliers to find that

$$\hat{p}_i = \frac{1}{n}\frac{1}{1+\lambda(\widehat{V}_i - \theta_{\mathbf{p}})}$$

where λ is the solution to

$$f(\lambda) = \sum_{i\in s} \frac{\widehat{V}_i - \theta_{\mathbf{p}}}{1+\lambda(\widehat{V}_i - \theta_{\mathbf{p}})} = 0.$$

After replacing $\hat{p}_i$ into (2.6.10) and applying the logarithm to $R(\theta)$ the profile jackknife EL ratio function is defined as

$$-2\log R(\theta) = \sum_{i\in s} \log\left\{1 + \lambda(\widehat{V}_i - \theta_{\mathbf{p}})\right\} \tag{2.6.11}$$

Jing et al. (2009) showed that (2.6.11) converges in distribution to a χ_1^2, as $n \to \infty$ and as consequence a $(1-\alpha)$ level confidence interval for θ is given $\mathcal{C}_j = \{|\theta| - 2\log R(\theta) \leqslant 0\}$ can be constructed. c is chosen to satisfy $P(\chi_1^2 \leqslant c) = 1-\alpha$. They showed that it works for the one and two-sample U-statistics but it is not possible to confirm that in general jackknife EL validity holds for any nonlinear statistic.

2.7 Bootstrap methods as alternative for building confidence intervals

The bootstrap method was introduced by Efron (1979) as a method to approximate estimation of variance and confidence intervals. Bootstrap techniques offer relatively simple options for deriving confidence intervals without rigorous assumptions about the distribution of data. A short description of the bootstrap technique is presented to facilitate the explanation of rescaled bootstrap (Rao and Wu, 1988) and direct bootstrap (Antal and Tillé, 2011) which are intended for variance estimation with complex survey data.

2.7.1 The Bootstrap

In the context of survey sampling the use of the bootstrap can be outlined with the next description: consider a probability sample s of size n selected from a population U given a sampling design without replacement. The population parameter θ is estimated by $\hat{\theta}$ and the aim is to estimate its variance $V(\hat{\theta})$. To compute $V(\hat{\theta})$ by simulation, s is assumed as an artificial population U^* that mimics the real, but unknown population U. Then, a series of independent samples or bootstrap samples are drawn with replacement from U^* using the same sampling design applied in the selection of s from U. For each bootstrap sample an estimator θ^*, $\{b = 1, \ldots, B\}$ is calculated in the same way as $\hat{\theta}$ was calculated. The empirical distribution of θ^* is considered as an estimate of the sampling distribution of $\hat{\theta}$ and $V(\hat{\theta})$ is estimated by

$$\widehat{V}[\widehat{\theta}]_{Boot} = \frac{1}{B-1}\sum_{b=1}^{B}(\hat{\theta}^{(b)} - \hat{\theta})^2 \qquad (2.7.1)$$

where $\widehat{\theta}$ could be replaced by $\theta^* = (1/B)\sum_{b=1}^{B}\hat{\theta}^*$. The bootstrap variance estimator $\widehat{V}[\widehat{\theta}]_{Boot}$ is not calculated when the main purpose is to obtain confidence intervals. Instead, the confidence intervals are defined by the observed distribution of θ^*. A $100(1-\ \alpha)\%$ confidence interval for θ is bounded by a pair of random numbers

$L = L(\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(B)})$ and $U = U(\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(B)})$ such that

$$\Pr[L \leqslant \theta \leqslant U] = 1 - \alpha.$$

For a 95% confidence intervals the resampled θ^* values are ordered getting $\theta^{(1)} \leqslant \theta^{(2)} \leqslant \cdots \leqslant \theta^{(B)}$. For large values of B about 95% of the resampled values are between $L = \theta^{(0.025B)}$ and $U = \theta^{(0.975B)}$, some rounding or interpolation is applied for non integer values of $0.025B$ and $0.975B$. Both bootstrap methods are concisely explained in the following subsections.

2.7.2 Rescaled Bootstrap

The rescaled bootstrap was proposed by Rao and Wu (1988) for variance estimation of non linear statistics $\widehat{\theta}$, expressed as functions of means, under stratified multistage designs and with replacement sampling of primary sampling units (PSU). The method allows the estimation of confidence intervals. Rao et al. (1992) modified the rescaled bootstrap to cover smooth and non smooth functions. They proposed to make the scale adjustment on the survey weights rather the data values. The method works as follows: for $b = 1, \ldots, B$ where B is a large number, draw a simple random sample of $n_h^{(*)}$ PSUs with replacement independently in each stratum h from the n_h sample PSUs. Let $m_{hi}^{(*)}$ the numbers of times the (hi)-th PSU is selected $\left(\sum_{(hi)\in s} m_{hi} = m_h\right)$. The bootstrap weights are defined as

$$w_{hik}^* = \left[\{1 - (m_h/\,(n_h - 1))\}^{1/2} + (m_h/\,(n_h - 1))^{1/2}\,(n_h/m_h)\,m_{hi}^*\right] w_{hik}$$

where w_{hik} are the initial sampling weights. $m_{hi}^* = 0$ if the (hi)-th PSU is not selected. Then, calculate the rescaled bootstrap estimator of θ^* using the normalized bootstrap weights w_{hik}^* in the formula for $\widehat{\theta}$. When $m_h^* = n_h - 1$ the formula for the rescaled weights becomes $w_{hik}^* = [m_h/\,(n_h - 1)]\,w_{hik}$. The method gives a consistent estimator of the variance of $\widehat{\theta}$, a smooth function of mean, which can be approximated by (2.7.1). If $n_h > 3$ and $m_h = (n_h - 2)^2 / (n_h - 1)$ this provides a bootstrap estimate distribution whose third moment matches the unbiased estimate of the third moment of the estimator, for the linear case. The same choice of m_h and when strata variances σ_h^2 are known, ensures the second-order term of Edgeworth expansion of $Z = (\widehat{\overline{Y}} - \overline{Y})/\sigma$ (σ^2 is the true variance of $\widehat{\overline{Y}}$) matches the second-order term of the bootstrap distribution of Z. Rao and Wu (1988) obtained a bootstrap-t confidence intervals for the same class of smooth functions, by approximating the distribution of $t = (\widehat{\theta} - \theta)/\hat{s}(\widehat{\theta})$ by its bootstrap counterpart $t^* = (\widehat{\theta}^* - \theta)/\hat{s}(\widehat{\theta}^*)$, where $\hat{s}(\widehat{\theta}^*)$ is approximate by jackknife variance estimation. The bootstrap confidence intervals are obtained from the histogram of t^*.

2.7.3 Direct bootstrap

Antal and Tillé (2011) proposed the direct bootstrap for variance estimation for samples selected with a complex design, such as simple random sampling with and without replacement, Poisson sampling, and unequal probability sampling with and without replacement. The method consists of generating a mixed sample by selecting a subsample of units without replacement and other subsample selected with replacement in order to adjust for the finite population setting. The method reproduces unbiased estimators of variance in the linear case and does not required rescaling, artificial populations or correction factors. The methodology uses two sampling designs for resampling the units, sampling with over-replacement (Antal and Tillé, 2010) and one-one sampling. A random sample selected with over-replacement is the vector $\underline{\boldsymbol{S}} = (S_1, \ldots, S_k, \ldots, S_N)^\mathsf{T} = (X_1, \ldots, X_k, \ldots, X_N)^\mathsf{T}$ such that $\sum_{k=1}^{N} X_k = n$, where S_k is the number of times the unit is selected in the sample, and X_k are geometric random variables such that $\Pr(X_k = x_k) = (1-p)\, p^{x_k},\ x_k = 0, 1, 2, \ldots$ and

$$\Pr(S_1 = x_1, \ldots, S_k = x_k, \ldots, S_N = x_N) = \binom{N+n-1}{n}^{-1}.$$

S_k has a negative hypergeometric distribution

$$\Pr(S_k = j) = \binom{N-1+n-j-1}{n-j}\binom{N+n-1}{n}^{-1}$$

$j = 0, \ldots, n$. For this sampling design the expectation of S_k is $\mathrm{E}\left(S_k\right) = n/N$ and the variance is

$$V(S_k) = n\left(N-1\right)\left(N+n\right)/N^2\left(N+1\right).$$

$V(\widehat{Y}) = N\left(N-1\right)\left(N+n\right)\left[n\left(N+1\right)\left(N-1\right)\right]^{-1}\sum_{k\in U}\left(y_k - \overline{Y}\right)^2$, Y is an estimator of a total. The one-one sampling is only a sampling design for resampling n units from a sample of size n with $\mathrm{E}\left(S_k\right) = 1$ and $V(S_k) = 1$, using a mixture of simple random sampling with replacement and sampling with over-replacement. This notation is used in the direct bootstrap method: $\mathrm{E}\left(S_k\right) = \pi_k$, $\mathrm{cov}\left(S_k, S_l\right) = \Delta_{kl}$, $\underline{\boldsymbol{S}}^* = \left(S_1^*, \ldots, S_k^*, \ldots, S_N^*\right)^{\mathsf{T}}$; S_k^* is the number of times the unit is resampled; $V(\widehat{Y}^*|\underline{\boldsymbol{S}})$ is the conditional variance of $\widehat{Y}^*$ given $\underline{\boldsymbol{S}}$; and $\mathrm{cov}\left(S_k^*, S_l^*|\underline{\boldsymbol{S}}\right) = \Sigma_{kl}$. The goal of the method is apply a resampling mechanism such that $\mathrm{E}\left(S_k^*|\underline{\boldsymbol{S}}\right) = 1$ holds in order to have $\mathrm{E}\left(Y^*|\underline{\boldsymbol{S}}\right) = Y$. $\Sigma_{kl} = \Delta_{kl}/\pi_{kl}$ is required to have $V(\widehat{Y}^*|\underline{\boldsymbol{S}}) = \widehat{\mathrm{V}}\left(\widehat{Y}\right)$. For samples selected with replacement and unequal inclusion probabilities the one-one sampling design can be used directly in the bootstrap method. In order to accomplish the correct variance if the samples are selected without replacement, a portion is resampled without replacement and another is selected according to a one-one design. The implementation of selecting resampled units consists of computing the samples sizes of the two components of the mixtures. Then select the sample sizes of the two components of mixture and then proceed to select the bootstrap samples. The direct bootstrap method proposed for a simple random sampling with replacement design and the one-one design as the resampling design generates $V(\widehat{Y}^*|\underline{\boldsymbol{S}}) = N^2\left[n\left(n-1\right)\right]^{-1}\sum_{k\in\underline{\boldsymbol{S}}}\left(y_k - \overline{Y}\right)^2$. In the case of resampling with the simple random sampling with replacement the variance is given by $V(\widehat{Y}^*|\underline{\boldsymbol{S}}) = N^2 n^{-2}\sum_{k\in\underline{\boldsymbol{S}}}\left(y_k - \overline{Y}\right)^2$. For the design of unequal probability sampling with replacement and one-one design as the replacement method the variance of $\widehat{Y}^*$ is $V(\widehat{Y}^*|\underline{\boldsymbol{S}}) = n\left(n-1\right)^{-1}\sum_{k\in\underline{\boldsymbol{S}}}\left(y_k/\pi_k - \widehat{Y}/n\right)^2$. The variance estimator of $\widehat{Y}^*$ for simple random sampling with replacement is

$$V(\widehat{Y}) = N\left(N-1\right)\left(N+n\right)\left[n\left(N+1\right)\left(N-1\right)\right]^{-1}\sum_{k\in\underline{\boldsymbol{S}}}\left(y_k - \overline{Y}\right)^2,$$

which is a result of a composite resampling design where the units of sample $\underline{\boldsymbol{S}}_C$ of size $m = n^2/N$ are selected with simple random sampling with replacement. The units of the sample $\underline{\boldsymbol{S}}_D$ are selected with the one-one design from the non-selected units in $\underline{\boldsymbol{S}}_C$. The bootstrap sample is $\underline{\boldsymbol{S}}^* = \underline{\boldsymbol{S}}_C + \underline{\boldsymbol{S}}_D$. Unequal probability sampling without replacement involves a more complex procedure. The direct bootstrap does not reproduce exactly its variance estimator but provides three approximations. The resampling design consists of two stages, in the first one a sample $\underline{\boldsymbol{S}}_C$ is selected from $\underline{\boldsymbol{S}}$ with unequal probability sampling without replacement and inclusion probabilities ϕ_k. In the second stage, a sample $\underline{\boldsymbol{S}}_D$ is selected with a one-one design from the non-selected unit in the first stage. The bootstrap sample is $\underline{\boldsymbol{S}}^* = \underline{\boldsymbol{S}}_C + \underline{\boldsymbol{S}}_D$. If

$$1 - \phi_k = \min[n\left(n-1\right)^{-1}\left(1-\pi_k\right)\left[1-\left(1-\pi_k\right)\left\{\sum_{j\in U} S_j\left(1-\pi_j\right)\right\}^{-1}\right], 1]$$

is chosen, the estimator of variance is close to the Hájek (1981) variance estimator $V(\widehat{Y}^*|\underline{\boldsymbol{S}}) \approx \widehat{\mathrm{V}}\left(\widehat{Y}\right) = \sum_{k\in\underline{\boldsymbol{S}}} c_k \left(\frac{y_k}{\pi_k} - \left(\sum_{k\in\underline{\boldsymbol{S}}} c_k\right) / \left(\sum_{k\in\underline{\boldsymbol{S}}} c_k \frac{y_k}{\pi_k}\right)\right)^2$, $c_k = n\left(n-1\right)^{-1}\left(1-\pi_k\right)$. When the choice is $\phi_k = \pi_k$ and c_k is such that $c_k - c_k^2 / \sum_{j\in U} S_j c_j = 1 - \pi_k$ the result is that $\widehat{\mathrm{V}}\left(\widehat{Y}\right)$ approximates to the variance estimator proposed by Deville and Tillé (2005). If the aim is reconstructing exactly the diagonal of Δ_{kl}/π_{kl}, then ϕ_k is selected such that $1 - \phi_k = \min\left[-\sum_{\substack{j\in U\\ j\neq k}} S_j \frac{\Delta_{kl}}{\pi_{kl}}, 1\right]$.

3

Empirical likelihood ratio estimator confidence intervals for unequal probability sampling

The first application of empirical likelihood in survey sampling was due to Chen and Qin (1993) for the use of auxiliary information under simple random sampling without replacement. The pseudo empirical likelihood was proposed for an unbiased estimation of the log likelihood function in the case of complex sampling designs. The function is based on an asymptotic χ^2 approximation to an adjusted pseudo empirical ratio function, which requires a scale factor to take into account the characteristics of the sampling design and the use of auxiliary information. The adjustment or scale factor is associated with the design effect; therefore, the estimation of variance and other source of randomness are included. The variance estimation involves the computation of second order inclusion probabilities that might be quite difficult, in particular when sampling designs have to be modified due to a set of constraints involving auxiliary information. The construction of pseudo empirical likelihood confidence intervals relies on variance estimates and design-effects

We propose an empirical likelihood ratio function that is asymptotically distributed as a χ^2 random variable and can be used to construct design-based confidence intervals for an unequal probability sampling design. It does not depend on variances estimates, design-effects, resampling or linearisation. This approach is convenient in applications when, for example the population distribution is highly skewed or the parameter of interest is not linear or the parameter estimator is biased. In this situation the assumption of normal distribution of the estimators may have low precision and as consequence the normal based confidence intervals are unsatisfactory. The proposed approach can be applied to the Horvitz and Thompson estimator, the Hájek estimator or regression estimator. We also suggest an empirical likelihood ratio function for a small and large sampling fraction. It also offers a likelihood-based justification for design-based approaches used in sample surveys.

3.1 Preliminaries

Let U be a finite population of N units, where N is fixed quantity which is not necessarily known. Suppose that the population parameter of interest θ_0 is the unique solution to the following equation (e.g. Qin and Lawless, 1994).

$$G(\theta) = 0, \quad \text{with} \quad G(\theta) = \sum_{i \in U} g_i(\theta), \tag{3.1.1}$$

where $g_i(\theta)$ is a function of θ and of the characteristics of the unit i. This function does not need to be differentiable. For simplicity $g_i(\theta)$ and θ_0 are considered as scalars but they can be vectors. For example, θ_0 is the population mean $\mu = N^{-1} \sum_{i \in U} y_i$, when $g_i(\theta) = y_i - \theta$, where the y_i are the values of a variable of interest. Other examples are ratios, low income measures, regression coefficients and M-estimators (Qin and Lawless, 1994; Binder and Kovacevìc, 1995; Deville, 1999). The aim of this chapter is to derive an empirical likelihood confidence interval for θ_0.

Suppose that we wish to estimate θ_0 from the data of a sample s of size n selected with a single stage unequal probabilities sample design. We consider that sample size n

is a fixed (non-random) quantity. We consider three cases:

1. Sampling with replacement (probability proportional-to-size sampling, or PPS sampling), where n denotes the number of draws and the sample is the set of n observations (already-drawn elements can be reselected).

2. Sampling without replacement (probability proportional-to-size sampling or πPS sampling).

3. Rao-Hartley-Cochran sampling design (Rao et al., 1962).

We adopt a non-parametric design-based approach, where the sampling distribution is specified by the sampling design and the values of the variable of interest are fixed (non-random) quantities.

Under the design-based approach, the standard likelihood is flat and cannot be used for inference (Godambe, 1966). Hartley and Rao (1968) introduced an empirical likelihood-based approach which does not rely on models. Owen (1988) brought this approach in the main stream statistics (see also Owen, 2001). Since Chen and Qin (1993) suggested its first application in survey sampling, there have been many recent developments of empirical likelihood based methods in survey sampling (e.g. Rao and Wu, 2009) and adaptive sampling (Salehi et al., 2010).

Standard confidence intervals based upon the central limit theorem can perform poorly when the sampling distribution is not normal. For example, the lower bound of a confidence interval can be negative even when the parameter of interest is positive. The coverages and the tail errors can be also different from their intended levels. On the other hand, empirical likelihood confidence intervals may be better in this situation, as empirical likelihood intervals are determined by the distribution of the data (Rao and Wu, 2009) and the range of the parameter space is preserved. Note that the empirical likelihood confidence intervals have better coverage when the variable of interest is skewed or contains many zeros (Chen et al., 2003) which is common in many surveys and with estimations of domains.

Chen and Sitter (1999) proposed a pseudo empirical likelihood approach which can be used to construct confidence intervals (Wu and Rao, 2006). The pseudo empirical log-likelihood ratio function depends on a population parameter (the design effect) which needs to be estimated, incurring an additional variability which may be affect the coverage of the confidence intervals. Variance estimates can also be instable with skewed data. Kim (2009) proposed an empirical likelihood approach function under Poisson sampling (see also Owen, 2001, Ch 6.), although it was not shown how to use this function to construct confidence intervals.

We propose to use an empirical likelihood approach which is different from the pseudo empirical likelihood approach. We show that the proposed empirical likelihood estimator is asymptotically equivalent to an optimal regression estimator (Montanari, 1987), when the parameter of interest is a total or mean. Wu and Rao (2006) proposed a more efficient pseudo empirical likelihood approach (EL2) when the variable of interest is correlated with the inclusion probabilities. However, this approach cannot be used to estimate totals and counts (e.g. cross-tabulation of categorical variables) when N is unknown, which is a common situation in social household surveys. The proposed approach does not rely on variance estimates, or population parameters to compute confidence intervals for totals or counts.

3.2 Empirical likelihood approach under unequal probability sampling

We propose to use the following empirical log-likelihood function

$$m = \sum_{i=1}^{n} \log(m_i), \tag{3.2.1}$$

where m_i is the unit mass of unit i in the population (e.g. Deville, 1999), which are such that a set of constraints, described below, always hold. Hartley and Rao (1969) showed that function (3.2.1) is a log-empirical likelihood function under PPS sampling with replacement, as m_i/N is the probability of observe the unit i. We propose to use (3.2.1), under a πPS without replacement sampling design (with fixed sample size), despite of units not being selected independently. The aim is to show that function (3.2.1) can be used for point estimation, testing statistical hypotheses and to construct confidence intervals. In contrast, the pseudo empirical likelihood approach is based on Kullback-Leibler distance function (Rao and Wu, 2009).

The maximum likelihood estimators of m_i are the values $\widehat{m}_i$ which maximise the *log-empirical likelihood function* $\ell(\widehat{m}) = \log(L(m))$ subject to the constraints $m_i \geq 0$ and

$$\sum_{i=1}^{n} m_i \boldsymbol{c}_i = \boldsymbol{C} \tag{3.2.2}$$

where $\boldsymbol{c}_i$ is a know $Q \times 1$ vector associated with the i-th sampled unit and $\boldsymbol{C}$ is a known $Q \times 1$ vector. Note that the $\boldsymbol{c}_i$ and $\boldsymbol{C}$ cannot be any vectors, as they must obey the regularity conditions (3.3.3)-(3.3.8), given in Section 3.3. The regularity conditions hold when all the components of $\boldsymbol{c}_i$ are bounded in probability. We assume $\boldsymbol{c}_i$ such that the matrix (3.2.7) is of full rank. Under sampling without replacement (πPS sampling) π_i denotes the inclusion probability of unit i. On the other hand, in the case of sampling with replacement (PPS sampling), $\pi_i = n\dot{\pi}_i$, where $\dot{\pi}_i$ is the probability that unit i is selected on the first draw, or the second draw, or any other given draw (Hansen and Hurwitz, 1943). It is assumed vector $\boldsymbol{c}_i$ includes the inclusion probabilities π_i. This means that there exists a known vector $\boldsymbol{t}$ such that $\boldsymbol{t}^\mathsf{T}\boldsymbol{c}_i = \pi_i$ and $\boldsymbol{t}^\mathsf{T}\boldsymbol{C} = \sum_{i \in U} \pi_i$. Indeed, in this case, we have that (3.2.2) implies that $\sum_{i=1}^{n} m_i \boldsymbol{t}^\mathsf{T}\boldsymbol{c}_i = \boldsymbol{t}^\mathsf{T}\boldsymbol{C}$ or equivalently

$$\sum_{i=1}^{n} m_i \pi_i = \sum_{i \in U} \pi_i. \tag{3.2.3}$$

The constraint (3.2.2) is such that the constraint (3.2.3) always holds. For example, when we have a single stratum, π_i could be the first component of the $\boldsymbol{c}_i$, in this case $\boldsymbol{t} = (1, 0, \ldots, 0)^\mathsf{T}$.

Because the sample size is fixed, the constraint (3.2.3) reduces to $\sum_{i=1}^{n} m_i \pi_i = n$. Thus, Equation (3.2.3) represents a fixed sample size design constraint. This constraint has been suggested by Wu and Rao (2006). Under equal probability sampling design with $\pi_i = n/N$ the constraint (3.2.3) reduces $\sum_{i=1}^{n} m_i = N$ which is a constraint adopted under equal probability sampling (e.g. Rao and Wu, 2009). We do not impose that $\sum_{i=1}^{n} m_i = N$ always holds, except when $\pi_i = n/N$. If we want to impose that $\sum_{i=1}^{n} m_i = N$, it is required to include the additional constraint $\sum_{i=1}^{n} m_i x_i = N$ considering the auxiliary variable $x_i = 1$ (see Section 3.8).

Note that the vector $\boldsymbol{C}$ is not necessarily a vector of fixed quantities. Hence $\boldsymbol{C}$ can be fixed or random. Possible choices for $\boldsymbol{c}_i$ and $\boldsymbol{C}$ are discussed in Sections 3.4–3.9. The constraint (3.2.2) resembles the constraint used in calibration (e.g Huang and Fuller, 1978; Deville and Särndal, 1992). However, we will see in Sections 3.4–3.9 that $\boldsymbol{C}$ is not a vector of population totals of auxiliary variables. Using a duality argument, it can be shown that this minimisation of (3.2.1) under the constraints (3.2.2) problem has a unique solution (e.g. Chen et al., 2002) which can be found by using the following Lagrangian function

$$\boldsymbol{Q}(m, \boldsymbol{\eta}) = \sum_{i=1}^{n} \log(m_i) - (\boldsymbol{t} + \boldsymbol{\eta})^{\mathsf{T}}(\sum_{i=1}^{n} \boldsymbol{c}_i - \boldsymbol{C}). \tag{3.2.4}$$

The values m_i and $\boldsymbol{\eta}$ which minimise (3.2.4) are the solutions to the following set of equations

$$\begin{aligned} \frac{\partial \boldsymbol{Q}(m, \boldsymbol{\eta})}{\partial m_i} &= 0 \\ \frac{\partial \boldsymbol{Q}(m, \boldsymbol{\eta})}{\partial (\boldsymbol{t} + \boldsymbol{\eta})} &= 0. \end{aligned}$$

The solution is

$$\widehat{m}_i = [(\boldsymbol{t} + \boldsymbol{\eta})^{\mathsf{T}} \boldsymbol{c}_i]^{-1} = (\pi_i + \boldsymbol{\eta}^{\mathsf{T}} \boldsymbol{c}_i)^{-1} \tag{3.2.5}$$

as $\boldsymbol{t}^{\mathsf{T}} \boldsymbol{c}_i = \pi_i$. The parameter $\boldsymbol{\eta}$ is such that (3.2.2) holds. This parameter can be computed using an iterative Newton-Raphson procedure. Consider the following $Q \times 1$ vector function of $\boldsymbol{\eta}$, $\boldsymbol{f}(\boldsymbol{\eta}) = \sum_{i=1}^{n} \widehat{m}_i \boldsymbol{c}_i$. A Taylor approximation of $\boldsymbol{f}(\boldsymbol{\eta})$ in the neighbourhood of a initial guess $\boldsymbol{\eta}_0$ gives

$$\boldsymbol{\eta} \simeq \boldsymbol{\eta}_0 + \widehat{\boldsymbol{\Delta}}(\boldsymbol{\eta}_0)^{-1}(\boldsymbol{C} - \boldsymbol{f}(\boldsymbol{\eta}_0)), \tag{3.2.6}$$

as the constraint (3.2.3) can be re-written as $\boldsymbol{f}(\boldsymbol{\eta}) = \boldsymbol{C}$. The $Q \times Q$ matrix $\widehat{\boldsymbol{\Delta}}(\boldsymbol{\eta}_0)$ is the following gradient

$$\widehat{\boldsymbol{\Delta}}(\boldsymbol{\eta}) = \frac{\partial \boldsymbol{f}(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = -\sum_{i=1}^{n} \boldsymbol{c}_i \boldsymbol{c}_i^{\mathsf{T}} (\pi_i + \boldsymbol{\eta}^{\mathsf{T}} \boldsymbol{c}_i)^{-2}. \tag{3.2.7}$$

The recursive formula (3.2.6) can be used to compute $\boldsymbol{\eta}$. For the first iteration, we used $\boldsymbol{\eta}_0 = \mathbf{0}$ which gives a new approximation of $\boldsymbol{\eta}$ using (3.2.6). This approximation is used as a new value for $\boldsymbol{\eta}_0$ which is substituted into (3.2.6). We repeat this process until convergence. Note that it is not necessary to know N in order to compute $\boldsymbol{\eta}$ and $\widehat{m}_i$. With the purpose of ensuring the constraint $\widehat{m}_i > 0$ holds for all i, we used the modified Newton-Raphson procedure (Chen et al., 2002). The modification ensure that the concave function (3.2.1) converges to the global maximum. This method also is known as damped Newton's method and involves regulating the step size (Polyak, 1987, p. 63).

3.3 Maximum Empirical Likelihood Estimator of θ

The empirical log-likelihood ratio function is defined in this section. This function is minimised by the maximum empirical likelihood estimate of θ.

Let $\widehat{m}_i$ be the values which maximise (3.2.1) subject to the constraints $\widehat{m}_i \geq 0$ and (3.2.2) for a given $\boldsymbol{c}_i$ and $\boldsymbol{C}$. The maximum value of the empirical log-likelihood function is given by $\ell(\widehat{m}) = \sum_{i=1}^{n} \log(\widehat{m}_i)$. Let $\widehat{m}_i^*(\theta)$ be the values which maximise (3.2.1) subject to the constraints $\widehat{m}_i \geq 0$ and (3.2.2) when $\boldsymbol{c}_i = \boldsymbol{c}_i^*$ and $\boldsymbol{C} = \boldsymbol{C}^*$, where $\boldsymbol{c}_i^* = (\boldsymbol{c}_i, g_i(\theta))^{\mathsf{T}}$ and $\boldsymbol{C}^* = (\boldsymbol{C}, 0)^{\mathsf{T}}$. Let $\ell(\widehat{m}^*, \theta) = \sum_{i=1}^{n} \log(\widehat{m}_i^*(\theta))$ be the maximum value of the empirical log-likelihood function. The *empirical log-likelihood ratio function* (or profile likelihood) is defined by the following function of θ.

$$\widehat{r}(\theta) = 2\left\{\ell(\widehat{m}) - \ell(\widehat{m}^*, \theta)\right\}. \tag{3.3.1}$$

The Function (3.3.1) is an empirical log-likelihood ratio function conditional on the variable $\boldsymbol{c}_i$ given that the Equation (3.2.4) is used for finding the maximum values of

$\ell(\widehat{m})$ and $\ell(\widehat{m}^*, \theta)$ (Owen (2001, Chapter 3.10), Qin and Lawless (1994)). The maximum empirical likelihood estimator $\widehat{\theta}$ of θ_0 is defined by the value of θ which minimises $\widehat{r}(\theta)$. As $\widehat{r}(\theta)$ is a positive function with a minimum value of zero, $\widehat{\theta}$ is the solution when $\widehat{r}(\theta) = 0$. This implies that $\widehat{\theta}$ is the solution of the following estimating equation

$$\widehat{G}(\theta) = 0, \quad \text{with} \quad \widehat{G}(\theta) = \sum_{i=1}^{n} \widehat{m}_i g_i(\theta), \tag{3.3.2}$$

where $\widehat{m}_i$ is defined by (3.2.5). We assume that $g_i(\theta)$ are such that $\widehat{G}(\theta) = 0$ has a unique solution. Note that when $\boldsymbol{c}_i = Nn^{-1}\pi_i$ and $\boldsymbol{C} = N$ (or equivalently $\boldsymbol{c}_i = \pi_i$ and $\boldsymbol{C} = n$), we have that $\boldsymbol{\eta} = 0$ and $\widehat{m}_i = \pi_i^{-1}$. Under πPS sampling $\widehat{G}(\theta) = \sum_{i=1}^{n} g_i(\theta)\pi_i^{-1}$. In this case $\widehat{\theta}$ is the Horvitz and Thompson (1952) estimator $\widehat{Y}_\pi = \sum_{i=1}^{n} y_i\pi_i^{-1}$ when $g_i(\theta) = y_i - n^{-1}\theta\pi_i$ (the estimator Hansen and Hurwitz (1943) is obtained under PPS sampling). When $g_i(\theta) = y_i - \theta N^{-1}$, the estimator $\widehat{\theta}_\pi$ is the Hájek (1971) ratio estimator $\widehat{Y}_H = N\widehat{N}_\pi^{-1}\widehat{Y}_\pi$, where $\widehat{N}_\pi = \sum_{i=1}^{n} \pi_i^{-1}$. The estimator $\widehat{Y}_H$ may not be as efficient as $\widehat{Y}_\pi$ when y_i and π_i are correlated (Rao, 1966), which may be the case, for example, with business surveys. Using the pseudo EL1 and EL2 it is not possible to obtain $\widehat{Y}_\pi$ (Wu and Rao, 2006).

In order to derive asymptotic properties of the proposed empirical likelihood approach, we assume the asymptotic framework proposed by Isaki and Fuller (1982), where $n \to \infty$ and $N \to \infty$. We consider that n/N does not necessarily tend to zero. The standard empirical likelihood approach (Owen, 1988; Kim, 2009) assumes that the sampling fraction is negligible ($n/N \to 0$). However, many surveys (business surveys) use sampling fractions which are not necessarily negligible. The proposed empirical likelihood approach does not rely on this assumption. The stochastic $O(\cdot)$, $o(\cdot)$, $O_p(\cdot)$ and $o_p(\cdot)$ are defined according to this asymptotic framework, where the convergence in probability is with respect to the sampling design (Isaki and Fuller, 1982).

Consider the following regularity conditions

$$\begin{aligned} nN^{-1}\pi_i^{-1} &= O(1) \text{ and } n^{-1}N\pi_i = O(1), \text{ for all } i \in s && (3.3.3)\\ N^{-1}\,\|\widehat{\boldsymbol{C}}_\pi - \boldsymbol{C}\| &= O_p(n^{-\frac{1}{2}}), && (3.3.4)\\ \max\{\|\boldsymbol{c}_i\|\colon i \in s\} &= o_p(n^{\frac{1}{2}}), && (3.3.5)\\ \|\widehat{\boldsymbol{S}}\| &= O_p(1), && (3.3.6)\\ \|\widehat{\boldsymbol{S}}^{-1}\| &= O_p(1), && (3.3.7)\\ \frac{1}{nN^\tau}\sum_{i=1}^{n}\frac{\|\boldsymbol{c}_i\|^\tau}{\pi_i^\tau} &= O_p(n^{-\tau}), && (3.3.8) \end{aligned}$$

where $\tau \leq 3$, with

$$\widehat{\boldsymbol{S}} = \frac{n}{N^2}\widehat{\boldsymbol{\Delta}}(\boldsymbol{0}) = -\frac{n}{N^2}\sum_{i=1}^{n}\frac{1}{\pi_i^2}\boldsymbol{c}_i\boldsymbol{c}_i^\top, \quad \text{and} \quad \widehat{\boldsymbol{C}}_\pi = \sum_{i=1}^{n}\frac{\boldsymbol{c}_i}{\pi_i}. \qquad (3.3.9)$$

The matrix $\widehat{\boldsymbol{\Delta}}(\boldsymbol{0})$ is given by (3.2.7) with $\boldsymbol{\eta} = \boldsymbol{0}$. The quantity $\|\boldsymbol{A}\|=$ trace$(\boldsymbol{A}^\top\boldsymbol{A})^{1/2}$ denotes the Euclidean norm.

The condition (3.3.4) holds when the central limit holds. For unequal probability sampling designs, Isaki and Fuller (1982) gave the conditions under which (3.3.4) holds (see also Krewski and Rao, 1981, p.1014). The condition (3.3.3) was proposed by Krewski and Rao (1981, p. 1014) (see also Kim, 2009). Chen and Sitter (1999, Appendix 2) showed that the condition (3.3.5) holds for common unequal probability samplings designs. We considered that the $\boldsymbol{c}_i$ are such that (3.3.6) and (3.3.7) hold. Note that these conditions hold when $\widehat{\boldsymbol{S}}$ is positive definite and there exists a positive definite matrix $\boldsymbol{S}$ such that $\|\widehat{\boldsymbol{S}} - \boldsymbol{S}\|= o_p(1)$. For example, the conditions (3.3.6) and (3.3.7) hold when $\boldsymbol{c}_i = Nn^{-1}\pi_i$. However, the conditions (3.3.7) does not hold when $\boldsymbol{c}_i = \pi_i$. Note that the $\widehat{m}_i$ obtained from $\boldsymbol{c}_i = Nn^{-1}\pi_i$ and $\boldsymbol{C} = N$ are the same as those obtained with $\boldsymbol{c}_i = \pi_i$ and $\boldsymbol{C} = n$, as the constraint is the same. We will see that some of the component of $\boldsymbol{c}_i$ have to be multiplied by Nn^{-1} for the conditions (3.3.6) and (3.3.7) to hold. The condition (3.3.8) is a Lyapunov-type condition for the existence of moments (e.g. Krewski and Rao 1981, p.1014; Deville and Särndal 1992, p. 381).

Theorem 3.3.1. *Under the conditions* (3.3.3)–(3.3.8)*, and when θ is such that*

$$\frac{1}{nN^2}\sum_{i=1}^{n}\breve{g}_i(\theta)^2 = O_p(n^{-2}) \tag{3.3.10}$$

for all θ

$$\widehat{G}(\theta) = \widehat{G}_\pi(\theta) + \widehat{\boldsymbol{B}}^\mathsf{T}(\boldsymbol{C} - \widehat{\boldsymbol{C}}_\pi) + o_p(Nn^{-\frac{1}{2}}), \tag{3.3.11}$$

$$\widehat{G}_\pi(\theta) = \sum_{i=1}^{n}\breve{g}_i(\theta) \tag{3.3.12}$$

where $\breve{g}_i(\theta) = g_i(\theta)\pi_i^{-1}$ and $\widehat{\boldsymbol{B}}$ is a vector of regression coefficients defined as

$$\widehat{\boldsymbol{B}} = \left(\sum_{i=1}^{n}\frac{1}{\pi_i^2}\boldsymbol{c}_i\boldsymbol{c}_i^\mathsf{T}\right)^{-1}\sum_{i=1}^{n}\frac{1}{\pi_i^2}g_i(\theta)\boldsymbol{c}_i. \tag{3.3.13}$$

Proof. Let $\widetilde{\boldsymbol{C}}_\pi = \widehat{\boldsymbol{C}}_\pi - \boldsymbol{C}$. Using Equation B.1.1 of Lemma B.1, $v_i = \pi_i^{-1}\boldsymbol{c}_i^\mathsf{T}\boldsymbol{\eta}$ and Lemma B.2, we have that

$$\begin{aligned}\widehat{G}(\theta) &= \widehat{G}_\pi(\theta) - \sum_{i=1}^{n}\frac{g_i(\theta)\boldsymbol{c}_i^\mathsf{T}\boldsymbol{\eta}}{\pi_i^2(1+v_i)} = \widehat{G}_\pi(\theta) + \frac{n}{N^2}\sum_{i=1}^{n}\frac{g_i(\theta)\boldsymbol{c}_i^\mathsf{T}}{\pi_i^2(1+v_i)}\widehat{\boldsymbol{S}}^{-1}\widetilde{\boldsymbol{C}}_\pi - \widehat{\boldsymbol{e}}_1 \\ &= \widehat{G}_\pi(\theta) + \frac{n}{N^2}\sum_{i=1}^{n}\frac{g_i(\theta)\boldsymbol{c}_i^\mathsf{T}}{\pi_i^2}\widehat{\boldsymbol{S}}^{-1}\widetilde{\boldsymbol{C}}_\pi - \widehat{\boldsymbol{e}}_1 - \widehat{\boldsymbol{e}}_2, \end{aligned} \tag{3.3.14}$$

where

$$\widehat{\boldsymbol{e}}_1 = nN^{-1}\sum_{i=1}^{n}\frac{g_i(\theta)\boldsymbol{c}_i^\mathsf{T}\widehat{\boldsymbol{e}}}{\pi_i^2(1+v_i)} \text{ and } \widehat{\boldsymbol{e}}_2 = nN^{-2}\sum_{i=1}^{n}\frac{g_i(\theta)\boldsymbol{c}_i^\mathsf{T}v_i}{\pi_i^2(1+v_i)}\widehat{\boldsymbol{S}}^{-1}\widetilde{\boldsymbol{C}}_\pi.$$

As $|1+v_i| > \gamma > 0$ (see proof of Lemma B.2), we have that

$$|\widehat{\boldsymbol{e}}_1| \le \frac{n}{N\gamma}\,\|\widehat{\boldsymbol{e}}\|\sum_{i=1}^{n}\frac{|\breve{g}_i(\theta)|\;\|\boldsymbol{c}_i\|}{\pi_i} \le \Big(\sum_{i=1}^{n}\breve{g}_i(\theta)^2\Big)^{\frac{1}{2}}\Big(\sum_{i=1}^{n}\frac{\|\boldsymbol{c}_i\|^2}{\pi_i^2}\Big)^{\frac{1}{2}} = O_p(Nn^{-1}),$$

using Cauchy's inequality, Equations (3.3.8) and (3.3.10). We have that

$$|\widehat{\boldsymbol{e}}_2| \le \frac{n}{N^2\gamma}\max[|v_i| : i \in s]\;\|\widehat{\boldsymbol{S}}^{-1}\|\|\widetilde{\boldsymbol{C}}_\pi\|\sum_{i=1}^{n}\frac{|\breve{g}_i(\theta)|\;\|\boldsymbol{c}_i\|}{\pi_i} = o_p(Nn^{-\frac{1}{2}}),$$

using Cauchy's inequality, Equations (3.3.4), (3.3.7), (3.3.8) and (3.3.10) and $\max[|v_i| : i \in s] = o_p(1)$. Thus, $|\widehat{\boldsymbol{e}}_1| + |\widehat{\boldsymbol{e}}_2| = o_p(Nn^{-\frac{1}{2}})$. Moreover, because of (3.3.9) the Equation (3.3.14) implies that $\widehat{G}(\theta) = \widehat{G}_\pi(\theta) - \big(\sum_{i=1}^n \pi_i^{-2}\boldsymbol{c}_i\boldsymbol{c}_i^\mathsf{T}\big)^{-1}\big(\sum_{i=1}^n \pi_i^{-2} g_i(\theta)\boldsymbol{c}_i^\mathsf{T}\big)\widetilde{\boldsymbol{C}}_\pi + o_p(Nn^{-\frac{1}{2}})$ which implies (3.3.11) because of (3.3.13). □

In Section 3.7 we give additional properties of (3.3.11).

3.4 Empirical likelihood confidence intervals

The main advantage of empirical likelihood approach is its capability of deriving non-parametric confidence intervals which do not depend on variance estimates. In this Section, we propose to use an empirical log-likelihood ratio function to derive empirical likelihood confidence intervals.

Empirical likelihood confidence intervals rely on the following conditions

$$\widehat{G}_\pi(\theta_0)V[\widehat{G}_\pi(\theta_0)]^{-\frac{1}{2}} \rightarrow N(0,1), \tag{3.4.1}$$

$$N^{-1}\widehat{G}_\pi(\theta_0) = O_p(n^{-\frac{1}{2}}), \tag{3.4.2}$$

$$\max\{|g_i(\theta_0)| : i \in s\} = o_p(n^{\frac{1}{2}}), \tag{3.4.3}$$

$$\frac{1}{nN^\tau}\sum_{i=1}^n |\breve{g}_i(\theta_0)|^\tau = O_p(n^{-\tau}), \tag{3.4.4}$$

where $V[\widehat{G}_\pi(\theta_0)]$ denotes the design based variance of $\widehat{G}_\pi(\theta_0)$ defined by (3.3.12) and $\tau \leqslant 3$. The conditions (3.4.2), (3.4.3) and (3.4.4) ensure that conditions (3.3.4)–(3.3.8) hold when $\boldsymbol{c}_i$ includes $g_i(\theta_0)$. Isaki and Fuller (1982) gave regularity conditions under which (3.4.2) holds. The condition (3.4.4) is a Lyapunov-type conditions for the existence of moments.

The condition (3.4.1) is weaker than the assumption of normality of $\widehat{\theta}$. As θ_0 is a constant, $\widehat{G}_\pi(\theta)$ is a Horvitz and Thompson (1952) estimator. Hájek (1964), Vísek (1979), Ohlsson (1986), Zhong and Rao (2000) and Berger (1998) gave regularity conditions for the asymptotic normality of the Horvitz and Thompson (1952) estimator. Under sampling with replacement, $\breve{g}_i(\theta_0)$ are independent, and standard large sampling theory can be used to show the normality (Prášková and Sen, 2009). Based on these evidences, it is reasonable to consider that for sampling with replacement design (3.4.1) holds, since $E[\widehat{G}_\pi(\theta_0)] = G(\theta_0) = 0$. Note that the classical empirical likelihood approach and the pseudo empirical likelihood also rely on (3.4.1) (e.g. Owen 1988, p. 242; Owen 2001, p. 219; Wu and Rao 2006, p. 364).

3.5 Empirical likelihood confidence intervals for sampling with replacement (PPS sampling)

Hartley and Rao (1969) showed that (3.2.1) is a log-empirical likelihood function under sampling with replacement (PPS sampling). Let $\widehat{m}_i$ be the values which maximise (3.2.1) subject to the constraints $\widehat{m}_i \geq 0$ and (3.2.2) for a given $\boldsymbol{c}_i$ and $\boldsymbol{C}$. The maximum value of the empirical log-likelihood function is given by $\ell(\widehat{m}) = \sum_{i=1}^{n} \log(\widehat{m}_i)$.

Corollary 3.5.1. *Under conditions* (3.3.4)-(3.3.8) *and* (3.4.1)-(3.4.4) *when* $\boldsymbol{c}_i = Nn^{-1}\pi_i$, $\boldsymbol{C} = N$, $\boldsymbol{c}_i^* = (Nn^{-1}\boldsymbol{c}_i^\top, g_i(\theta))^\top = (Nn^{-1}\pi_i, g_i(\theta))^\top$ *and* $\boldsymbol{C}^* = (Nn^{-1}\boldsymbol{C}^\top, 0)^\top = (N, 0)^\top$, *we have that*

$$\widehat{r}(\theta_0) = \widehat{G}_\pi(\theta_0)^2 \widehat{V}_{\text{PPS}}[\widehat{G}_\pi(\theta_0)]^{-1} + o_p(1) \tag{3.5.1}$$

where θ_0 *denotes the population parameter to estimate and* $\widehat{V}_{\text{PPS}}[\widehat{G}_\pi(\theta_0)]$ *is the following* PPS *variance estimator*

$$\widehat{V}_{\text{PPS}}[\widehat{G}_\pi(\theta_0)] = \sum_{i=1}^{n} \left(\breve{g}_i(\theta_0) - n^{-1}\widehat{G}_\pi(\theta_0)\right)^2, \tag{3.5.2}$$

where $\breve{g}_i(\theta_0) = g_i(\theta_0)\pi_i^{-1}$.

Proof. Suppose we have one stratum ($H = 1$). We obtain this corollary by substituting $\boldsymbol{z}_i$ by π_i in Corollary 3.8.2. In this case (3.6.2) reduces to (3.5.2). □

Note that we obtain the same $\widehat{m}_i$ and $\widehat{m}_i^*(\theta)$, when $\boldsymbol{c}_i = \pi_i$ and $\boldsymbol{C} = n$. Therefore, N does not need to be known. Under sampling with replacement with unequal probabilities (PPS sampling), the estimator (3.5.2) is a consistent estimator for the variance (e.g. Durbin, 1953; Särndal et al., 1992, p. 99). Hence the property (3.4.1) implies that $\widehat{r}(\theta_0)$ follows asymptotically a chi-squared distribution with one degree of freedom (Slutsky's theorem). Thus, the $(1-\alpha)$ level empirical likelihood confidence intervals (Wilks 1938; Hudson 1971) for the population parameter θ_0 is given by

$$\left\{\theta : \widehat{r}(\theta) \leq \chi_1^2(\alpha)\right\} = \left[\min\left\{\theta|\widehat{r}(\theta) \leq \chi_1^2(\alpha)\right\}; \max\left\{\theta|\widehat{r}(\theta) \leq \chi_1^2(\alpha)\right\}\right], \quad (3.5.3)$$

where $\chi_1^2(\alpha)$ is the upper α-quantile of the chi-squared distribution with one degree of freedom. Note that $\widehat{r}(\theta)$ is a convex non-symmetric function with a minimum when θ is the maximum empirical likelihood estimator. The convexity of $\widehat{r}(\theta)$ is verified by noting that $\ell(\widehat{m})$ does not involve θ and $\ell(\widehat{m}^*, \theta) = \sum_{i=1}^n \log\left(\widehat{m}_i^*(\theta)\right)$ is a concave function of θ. Then $-2 \times \ell(\widehat{m}^*, \theta)$ is a convex function of θ. This interval can be found using a bisection method (Chen and Qin, 2003; Wu, 2005). This involves calculating $\widehat{r}(\theta_0)$ for several values of θ. Note that (3.5.3) will give confidence intervals with the right coverage even when $\widehat{\theta}$ is biased. If $g_i(\theta_0)$ is a q-vector, the random variable $\widehat{r}(\theta_0)$ will converge to a chi-squared distribution with q degrees of freedom.

If we want to test a statistical hypothesis $H_0 : \theta_0 = \theta_0^\bullet$ versus $H_1 : \theta_0 \neq \theta_0^\bullet$, the p-value is given by $\Pr\left\{\chi_1^2 \geqslant \widehat{r}(\theta_0^\bullet)\right\}$.

3.6 Empirical likelihood Approach Under Stratified PPS Sampling with Replacement

Assume that the sample s is randomly selected by a one-stage stratified probability sampling design $p(s)$. Suppose that the finite population U is stratified into H strata denoted by $U_1, \ldots, U_h, \ldots, U_H$, where $\bigcup_{h=1}^H U_h = U$. Suppose that a sample s_h of a fixed size n_h is selected without replacement with unequal probabilities π_i from U_h. We assume that the number of strata H is bounded.

The empirical likelihood estimator is still the solution of (3.3.2) where $\widehat{m}_i$ are the values which maximise (3.2.1) under a set of constraints (3.2.4). The variables of the design (or stratification) are include in the constraints (3.2.4) as $\boldsymbol{c}_i = \boldsymbol{z}_i$ and $\boldsymbol{C} = \boldsymbol{n}$, where

$$\boldsymbol{z}_i = (z_{i1}, \ldots, z_{iH})^\top \quad \text{and} \quad \boldsymbol{n} = (n_1, \ldots, n_H)^\top \tag{3.6.1}$$

denotes the vector of the stratum sample size, with $z_{ih} = \pi_i$ when $i \in U_h$ and $z_{ih} = 0$ otherwise. It can be shown that $\widehat{m}_i = \pi_i^{-1}$.

Corollary 3.6.1. *Let $\boldsymbol{c}_i = Nn^{-1}\boldsymbol{z}_i$, $\boldsymbol{C} = Nn^{-1}\boldsymbol{n}$, $\boldsymbol{c}_i^* = (Nn^{-1}\boldsymbol{z}_i^\top, g_i(\theta))^\top$ and $\boldsymbol{C}^* = (Nn^{-1}\boldsymbol{n}^\top, 0)^\top$. We have that* (3.5.1) *holds where $\widehat{V}_{\text{PPS}}[\widehat{G}_\pi(\theta_0)]$ is now the stratified variance* PPS *estimator*

$$\widehat{V}_{\text{ST}}[\widehat{G}_\pi(\theta_0)] = \sum_{h=1}^{H} \left[\sum_{i \in s_h} \left(\breve{g}_i(\theta_0) - n_h^{-1} \widehat{G}_{\pi;h}(\theta_0) \right)^2 \right], \tag{3.6.2}$$

where $\widehat{G}_{\pi;h}(\theta_0) = \sum_{i \in s_h} \breve{g}_i(\theta_0)$.

Proof. This corollary is obtained by using a similar proof of Theorem B.5 and Corollary 3.8.2 with $\psi_i = \psi_i^\bullet = 1$. □

This variance estimator is consistent because the number of strata is bounded. Hence $\widehat{r}(\theta_0)$ follows a chi-squared distribution asymptotically and the empirical likelihood confidence intervals (3.5.3) can be computed with (3.3.1).

Note that we propose to use the same likelihood function (3.2.3) with or without stratification. With pseudo empirical likelihood approach, the pseudo empirical likelihood function without stratification is different from pseudo empirical likelihood function with stratification (e.g. Rao and Wu, 2009, p. 195)

3.7 Empirical Likelihood Approach with Auxiliary Variables

Let $\boldsymbol{x}_i$ be a P-component vector of auxiliary variables attached to the unit i. Let $\boldsymbol{f}_i\left(\boldsymbol{x}_i, \boldsymbol{X}\right) = \boldsymbol{x}_i - \boldsymbol{X}\pi_i n^{-1}$ a P-component vector. These variables are such that their population control totals $\boldsymbol{X} = \sum_{i\in U} \boldsymbol{x}_i$ are known. Let $\widehat{m}_i(\boldsymbol{x}_i)$ be the values which maximise (3.2.2) under the constraint (3.2.3) with $\boldsymbol{c}_i = \left(Nn^{-1}\boldsymbol{z}_i^{\mathsf{T}}, \boldsymbol{f}_i(\boldsymbol{x}_i, \boldsymbol{X})^{\mathsf{T}}\right)^{\mathsf{T}}$ and $\boldsymbol{C} = \sum_{i\in U} \boldsymbol{c}_i = (Nn^{-1}\boldsymbol{n}^{\mathsf{T}}, \boldsymbol{0})^{\mathsf{T}}$. The maximum empirical likelihood estimator is the solution of $\sum_{i=1}^{n} \widehat{m}_i(\boldsymbol{x}_i) g_i(\theta) = 0$ (see (3.3.2) and (3.3.3)). The $\widehat{m}_i(\boldsymbol{x}_i)$ are calibrated weights because $\sum_{i=1}^{n} \widehat{m}_i(\boldsymbol{x}_i \boldsymbol{f}_i\left(\boldsymbol{x}_i, \boldsymbol{X}\right)) = \boldsymbol{0}$ implies that $\sum_{i=1}^{n} \widehat{m}_i(\boldsymbol{x}_i) = \boldsymbol{X}$. In Theorem 3.7.1 the conditions under which empirical log-likelihood ratio function $\widehat{r}(\theta_0)$ follows asymptotically a chi-squared distribution with one degree of freedom are given, assuming that regression estimator (3.3.12) has a normal distribution asymptotically. Evidence of normality of the regression estimator are given in Scott and Wu (1981).

Theorem 3.7.1. *Let* $\boldsymbol{c}_i = \psi_i(Nn^{-1}\boldsymbol{z}_i^{\mathsf{T}}, \boldsymbol{f}_i^{\mathsf{T}})^{\mathsf{T}}$ *and* $\boldsymbol{C} = (Nn^{-1}\sum_{i=1}^{n} \psi_i \breve{\boldsymbol{z}}_i^{\mathsf{T}}, N^{-1}\sum_{i=1}^{n}(\psi_i - 1)\breve{\boldsymbol{f}}_i^{\mathsf{T}})^{\mathsf{T}}$, *where* $\breve{\boldsymbol{f}}_i = \boldsymbol{f}_i\pi_i^{-1}$ *and* $\boldsymbol{z}_i$ *is the* $H-$*vector of stratification and* $\boldsymbol{f}_i$ *is a* $P-$*vector. Let* $\ell(\widehat{m}) = \sum_{i=1}^{n} \log(\widehat{m}_i)$, *where* $\widehat{m}_i$ *is defined by* (3.2.5), *and* $\boldsymbol{\eta}$ *is such that* $\sum_{i=1}^{n} \widehat{m}_i \boldsymbol{c}_i = \boldsymbol{C}$. *Let* $\boldsymbol{c}_i^* = (\boldsymbol{c}_i^{\mathsf{T}}, \psi_i g_i(\theta_0))^{\mathsf{T}}$ *and* $\boldsymbol{C}^* = (\boldsymbol{C}^{\mathsf{T}}, \sum_{i=1}^{n}(\psi_i - 1)\breve{g}_i(\theta_0))^{\mathsf{T}}$. *Let* $\ell(\widehat{m}^*, \theta_0) = \sum_{i=1}^{n} \log(\widehat{m}_i^*\left(\theta_0\right))$, *where* $\widehat{m}_i^*\left(\theta\right)$ *is defined by* $\widehat{m}_i^*\left(\theta\right) = \left(\pi_i + {\boldsymbol{\eta}^*}^{\mathsf{T}} \boldsymbol{c}_i^*\right)^{-1}$, *and* $\boldsymbol{\eta}^*$ *is such that* $\sum_{i=1}^{n} \widehat{m}_i^* \boldsymbol{c}_i^* = \boldsymbol{C}^*$. *The values* ψ_i *are such that the regularity conditions* (3.3.4)-(3.3.8) *hold for* $\boldsymbol{c}_i = \boldsymbol{c}_i^*$, $\boldsymbol{C} = \boldsymbol{C}^*$. *We have that*

$$\widehat{r}(\theta_0) = 2\left\{\ell(\widehat{m}) - \ell(\widehat{m}^*, \theta_0)\right\} = \widehat{G}_{\text{REG}}(\theta_0)^2 \widehat{V}[\widehat{G}_{\text{REG}}(\theta_0)]^{-1} + O_p(n^{-\frac{1}{2}}) \qquad (3.7.1)$$

where $\widehat{G}_{\text{REG}}(\theta_0)$ *is the regression estimator defined by* (3.3.12) *with* $\boldsymbol{c}_i$ *and* $\boldsymbol{C}$ *defined above, where* $\widehat{V}[\widehat{G}_{\text{REG}}(\theta_0)] = \widetilde{\sigma}_{gg} - \widetilde{\boldsymbol{\Sigma}}_{zg}^{*\mathsf{T}} \widetilde{\boldsymbol{\Sigma}}_{zz}^{*-1} \widetilde{\boldsymbol{\Sigma}}_{zg}^{*}$,

$$\widetilde{\sigma}_{gg} = \sum_{i=1}^{n} \psi_i^2 \left(\frac{\widetilde{g}_i(\theta_0)}{\pi_i}\right)^2, \widetilde{\boldsymbol{\Sigma}}_{zz}^{*} = \frac{N^2}{n^2} \sum_{i=1}^{n} \psi_i^2 \breve{\boldsymbol{z}}_i^{\mathsf{T}} \breve{\boldsymbol{z}}_i, \quad \widetilde{\boldsymbol{\Sigma}}_{zg}^{*} = \frac{N}{n} \sum_{i=1}^{n} \psi_i^2 \breve{\boldsymbol{z}}_i \frac{\widetilde{g}_i(\theta_0)}{\pi_i}$$

where the $\widetilde{g}_i(\theta_0)$ *are the residuals defined by*

$$\widetilde{g}_i(\theta_0) = g_i(\theta_0) - \widehat{\boldsymbol{B}}^{\mathsf{T}}(N^{-1}\boldsymbol{z}_i^{\mathsf{T}}, \boldsymbol{f}_i^{\mathsf{T}})^{\mathsf{T}} \qquad (3.7.2)$$

and $\widehat{\boldsymbol{B}}$ *is defined by* (3.3.13).

Proof. Using Lemmas B.3 and B.4, we have that

$$-2[\ell(\widehat{m}_i)+\ell(\pi)] = (\widehat{\boldsymbol{C}}_\pi-\boldsymbol{C})^\mathsf{T}\widehat{\boldsymbol{\Sigma}}^{-1}(\widehat{\boldsymbol{C}}_\pi-\boldsymbol{C})+o_p(1), \tag{3.7.3}$$

where $\ell(\pi)=\sum_{i=1}^n \log(\pi_i)$ and $\widehat{\boldsymbol{\Sigma}}=\sum_{i=1}^n \pi_i^{-2}\boldsymbol{c}_i\boldsymbol{c}_i^\mathsf{T}$.

Let

$$\widehat{\boldsymbol{\Sigma}}^* = \sum_{i=1}^n \frac{1}{\pi_i}\boldsymbol{c}_i^*\boldsymbol{c}_i^{*\mathsf{T}} = \begin{pmatrix} \widehat{\boldsymbol{\Sigma}} & \widehat{\boldsymbol{\Sigma}}_{cg} \\ \widehat{\boldsymbol{\Sigma}}_{cg}^\mathsf{T} & \widehat{\sigma}_{gg}\end{pmatrix}, \tag{3.7.4}$$

where $\widehat{\boldsymbol{\Sigma}}_{cg}$ is the $(H+P)\times 1$ sub-matrix. Consider $\widetilde{\boldsymbol{c}}_i^* = \psi_i(N^{-1}\boldsymbol{z}_i^\mathsf{T}, \boldsymbol{f}_i^\mathsf{T}\widetilde{g}_i(\theta_0))^\mathsf{T}$, $\widetilde{\boldsymbol{C}}^* = (\boldsymbol{C}^\mathsf{T}, \sum_{i=1}^n(\psi_i-1)\widetilde{g}_i(\theta_0)-\widehat{\boldsymbol{B}}^\mathsf{T}\boldsymbol{C})^\mathsf{T}$, where $\widetilde{g}_i(\theta_0)$ are defined by (3.7.2). We have that

$$\widetilde{\boldsymbol{\Sigma}}^* = \sum_{i=1}^n \frac{1}{\pi_i^2}\widetilde{\boldsymbol{c}}_i^*\widetilde{\boldsymbol{c}}_i^{*\mathsf{T}} = \begin{pmatrix} \widehat{\boldsymbol{\Sigma}} & \boldsymbol{0} \\ \boldsymbol{0} & \widetilde{\sigma}_{gg}\end{pmatrix} \tag{3.7.5}$$

$$\widetilde{\boldsymbol{C}}_\pi^* - \widetilde{\boldsymbol{C}}^* = \sum_{i=1}^n \widetilde{\boldsymbol{c}}_i^*\pi_i^{-1} - \widetilde{\boldsymbol{C}}^* = ((\widehat{\boldsymbol{C}}_\pi-\boldsymbol{C})^\mathsf{T}, \widehat{G}_{\text{REG}}(\theta_0))^\mathsf{T}, \tag{3.7.6}$$

where $\widehat{G}_{\text{REG}}(\theta_0)$ is the regression estimator defined by (3.3.12) with the $\boldsymbol{c}_i$ and $\boldsymbol{C}$ considered. This estimator is given by $\widehat{G}_{\text{REG}}(\theta_0) = \sum_{i=1}^n \breve{g}_i(\theta_0) - \widehat{\boldsymbol{B}}^\mathsf{T}(\boldsymbol{0}_H, \widehat{\boldsymbol{f}}_\pi)^\mathsf{T} = \sum_{i=1}^n \widetilde{g}_i(\theta_0)$, with $\widehat{\boldsymbol{f}}_\pi = \sum_{i=1}^n \check{\boldsymbol{f}}_i$ and $\widehat{\boldsymbol{B}} = \widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{\Sigma}}_{cg}$.
The matrix $\widehat{\boldsymbol{\Sigma}}^*$ is a block diagonal matrix because its extra diagonal block is given by

$$\sum_{i=1}^n \frac{\boldsymbol{c}_i^\mathsf{T}\psi_i\widetilde{g}_i(\theta_0)}{\pi_i^2} = \sum_{i=1}^n \frac{\boldsymbol{c}_i^\mathsf{T}\psi_i g_i(\theta_0)}{\pi_i^2} - \sum_{i=1}^n \frac{\boldsymbol{c}_i\boldsymbol{c}_i^\mathsf{T}}{\pi_i^2}\widehat{\boldsymbol{B}} = \widehat{\boldsymbol{\Sigma}}_{cg} - \widehat{\boldsymbol{B}}^\mathsf{T}\widehat{\boldsymbol{\Sigma}} = \boldsymbol{0} \tag{3.7.7}$$

as $\widetilde{g}_i(\theta_0) = g_i(\theta_0) - \widehat{\boldsymbol{B}}^\mathsf{T}\boldsymbol{c}_i\psi_i^{-1}$ and $\widehat{\boldsymbol{B}} = \widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{\Sigma}}_{cg}$. Using Equations (3.7.5) and (3.7.6), we have that

$$(\widetilde{\boldsymbol{C}}_\pi^* - \widetilde{\boldsymbol{C}}^*)^\mathsf{T}\widehat{\boldsymbol{\Sigma}}^{*-1}(\widetilde{\boldsymbol{C}}_\pi^* - \widetilde{\boldsymbol{C}}^*) = (\widehat{\boldsymbol{C}}_\pi-\boldsymbol{C})^\mathsf{T}\widehat{\boldsymbol{\Sigma}}^{-1}(\widehat{\boldsymbol{C}}_\pi-\boldsymbol{C}) + \widehat{G}_{\text{REG}}(\theta_0)^2\widetilde{\sigma}_{gg}^{-1}. \tag{3.7.8}$$

Let $\widetilde{m}_i^*(\theta) = (\pi_i - \widetilde{\boldsymbol{\eta}}^{*\mathsf{T}}\widetilde{\boldsymbol{c}}_i)^{-1}$, where $\widetilde{\boldsymbol{\eta}}^*$ is such that $\sum_{i=1}^n \widetilde{m}_i^*(\theta)\,\widetilde{\boldsymbol{c}}_i^* = \widetilde{\boldsymbol{C}}_\pi^*$. As this constraint is just a linear transformation of the constraint $\sum_{i=1}^n \widehat{m}_i^*\boldsymbol{c}_i^* = \boldsymbol{C}^*$, we have that $\widetilde{m}_i^*(\theta) = \widehat{m}_i^*$ and $\sum_{i=1}^n \log(\widetilde{m}_i^*(\theta)) = \sum_{i=1}^n \log(\widehat{m}_i^*) = \ell(\widehat{m}^*, \theta)$. Using Lemmas

B.3 and B.4, we have that $-2[\ell(\widehat{m}_i)+\ell(\pi)] = (\widetilde{\boldsymbol{C}}^*_\pi - \widetilde{\boldsymbol{C}}^*)^\mathsf{T}\widehat{\boldsymbol{\Sigma}}^{*-1}(\widetilde{\boldsymbol{C}}^*_\pi - \widetilde{\boldsymbol{C}}^*) + O_p(n^{-\frac{1}{2}})$. Thus (3.7.3) and (3.7.8) imply that

$$\widehat{r}(\theta_0) = 2\left\{\ell(\widehat{m}) - \ell(\widehat{m}^*, \theta_0)\right\} = \widehat{G}_{\text{REG}}(\theta_0)^2 \widetilde{\sigma}_{gg}^{-1} + o_p(1). \tag{3.7.9}$$

Note that (3.7.7) implies $\widetilde{\boldsymbol{\Sigma}}^*_{zg} = \mathbf{0}_H$. Hence (3.7.9) implies (3.7.1). □

When $\psi_i = q_i = (1-\pi_i)^{1/2}$, the variance $\widehat{V}[\widehat{G}_{\text{REG}}(\theta_0)]$ takes into account the calibration constraint and the fixed sizes constraints. Deville and Tillé (2005) showed that $\widehat{V}[\widehat{G}_{\text{REG}}(\theta_0)]$ is a consistent variance estimator under stratified with high entropy designs. When $\psi_i = q_i$, the quantity $\widehat{V}[\widehat{G}_{\text{REG}}(\theta_0)]$ is the variance estimator under a stratified with replacement PPS sampling design. In Section 3.8 we describe a relevant property of q_i. The proposed approach allows for the calibration of parameters more complex than totals, for example quantiles, variances or means. In this case, the calibration constraint is specified by the following set of estimating equations $\sum_{i=1}^n m_i \boldsymbol{f}_i(\boldsymbol{x}_i, \boldsymbol{\vartheta}_0) = 0$, where $\boldsymbol{f}_i(\boldsymbol{x}_i, \boldsymbol{\vartheta}_0)$ is a vector function of the auxiliary variables and a known parameter $\boldsymbol{\vartheta}_0$. The approach described in this section can be used in this situation after substituting $\boldsymbol{f}_i(\boldsymbol{x}_i, \boldsymbol{X})$ by $\boldsymbol{f}_i(\boldsymbol{x}_i, \boldsymbol{\vartheta}_0)$. For example, $\boldsymbol{f}_i(\boldsymbol{x}_i, \boldsymbol{\vartheta}_0) = \boldsymbol{x}_i - \boldsymbol{\vartheta}_0$ is used when the calibration is with respect to known population means $\boldsymbol{\vartheta}_0 = XN^{-1}$.

Consider the particular case when we have a single auxiliary variable x_i with a known control total $X = \sum_{i\in U} x_i$. Let $\boldsymbol{c}_i = (Nn^{-1}\pi_i, f_i(x_i, X))^\mathsf{T}$ and $\boldsymbol{C} = (N, 0)^\mathsf{T}$, (or equivalently $\boldsymbol{c}_i = (\pi_i, f_i(x_i, X))^\mathsf{T}$ and $\boldsymbol{C} = (n, 0)^\mathsf{T}$) where $f_i(x_i, X) = x_i - X\pi_i n^{-1}$. Using (3.3.12), we have that

$$\widehat{G}(\theta) = \widehat{G}_\pi(\theta) + \widehat{B}_x\left(X - \widehat{X}_\pi\right) + o_p(N), \tag{3.7.10}$$

where $\widehat{X}_\pi = \sum_{i=1}^n \breve{x}_i$ and

$$\widehat{B}_x = \frac{\sum_{i=1}^n \left(\breve{x}_i - n^{-1}\widehat{X}_\pi\right)\left(\breve{g}_i(\theta) - n^{-1}\widehat{G}_\pi(\theta)\right)}{\sum_{i=1}^n \left(\breve{x}_i - n^{-1}\widehat{X}_\pi\right)^2}, \tag{3.7.11}$$

$\breve{x}_i = x_i\pi_i^{-1}$ and $\breve{g}_i(\theta) = g_i(\theta)\pi_i^{-1}$. Note that $\widehat{B}_x$ is the estimator of the covariance between $\widehat{G}_\pi(\theta)$ and $\widehat{X}_\pi$ divided by the estimator of the variance of $\widehat{X}_\pi$ under a with replacement PPS sampling design (e.g. Särndal et al., 1992, p. 89). Therefore $\widehat{B}_x$ is the optimal regression coefficient (e.g. Isaki and Fuller, 1982; Montanari, 1987; Rao, 1994; Särndal, 1996; Berger et al., 2003) when the sampling fraction is small. This results can be extended when we have more than one auxiliary variable. Hence, the empirical likelihood estimator is asymptotically optimal. When N is known, the efficiency of the maximum empirical likelihood estimator can be improved by setting $x_i = 1$ or including a variable equal to one into $\boldsymbol{x}_i$. For example, when $\boldsymbol{c}_i = (Nn^{-1}\pi_i, f_i)^\mathsf{T}$ and $\boldsymbol{C} = (N, 0)^\mathsf{T}$, where $f_i = 1 - \pi_i N n^{-1}$, Equation (3.7.10) implies that

$$\widehat{G}(\theta) = \widehat{G}_\pi(\theta) + \widehat{B}_x \left(\widehat{N}_\pi - N\right) nN^{-1} + o_p(N) \tag{3.7.12}$$

where $\widehat{N}_\pi = \sum_{i=1}^n \pi_i^{-1}$, and $\widehat{B}_x$ is given by (3.7.11) after substituting x_i by $\pi_i - nN^{-1}$. Equation (3.5) in Kim (2009) proposed an estimator optimal under Poisson sampling with negligible sampling fractions, by considering the constraints $\sum_{i=1}^n (1, x_i)^\mathsf{T} P_i = \sum_{i\in U}(1, x_i)^\mathsf{T}$, where P_i is a probability mass. If x_i is replaced by π_i in equation (3.5) in Kim (2009), we obtain the same constraints that gives (3.7.11). In this case, equation (3.5) in Kim (2009) reduces to a similar estimator: $\widehat{G}_\pi(\theta)+\widetilde{B}_x \left(\widehat{N}_\pi - N\right) nN^{-1}\widehat{N}_\pi N^{-1}$, where $\widetilde{B}_x$ is given by (3.7.11) after substituting $n^{-1}\widehat{G}_\pi(\theta)$ by $\widehat{G}_\pi(\theta)\widehat{N}_\pi^{-1}\pi_i^{-1}$ and x_i by $\pi_i - nN^{-1}$. Both regression coefficients are only approximately equal. The difference between both regression coefficients is due to Kim (2009)'s estimator being based on Poisson sampling design. Under simple random sampling, Hartley and Rao (1968) showed that in the same situation, $\widehat{G}(\theta)$ is asymptotically equivalent to the customary regression estimator (see also Chen and Qin, 1993; Owen, 2001, Chapter 8) with π_i as auxiliary variable. This customary estimator is different from (3.7.10), and inconsistent under unequal probability sampling. This issue is due to the fact Hartley and Rao (1968) result does not hold under unequal probability sampling. The approximation (3.7.10) is based upon a different set of regularity conditions which takes into account the unequal probabilities. Under simple random sampling, $G(\theta)$ reduces to the customary regression estimator which is consistent under simple random sampling.

3.8 Empirical likelihood confidence intervals for πPS without replacement

In business surveys it is common practice to have large sampling fractions n/N. For this case, the PPS variance estimator (3.5.3) is biased, which implies that the empirical log-likelihood ratio function (3.5.3) cannot be used for confidence intervals because this random variable does not necessarily follow a chi-squared distribution, and needs to be adjusted to allow large sampling fractions. Note that the point estimation given by the solution to (3.3.2) with $\boldsymbol{c}_i$ and $\boldsymbol{C}$ described in Section 3.3 is still valid even if we have large sampling fractions. In this section, we propose to adjust the constraints $\boldsymbol{c}_i$, $\boldsymbol{C}$, $\boldsymbol{c}_i^*$ and $\boldsymbol{C}^*$ under a single stage sampling design without replacement with large sampling fractions. The purpose of the adjustment is to maintain the asymptotically chi-squared distribution of (3.3.1) under this sampling design. When we have a single stratum, we propose to use $\boldsymbol{c}_i = Nn^{-1}\pi_i$, $\boldsymbol{C} = N$, $\boldsymbol{c}_i^* = q_i(Nn^{-1}\pi_i, g_i(\theta))^\top$ and $\boldsymbol{C}^* = \left(\sum_{i=1}^n Nn^{-1}q_i, \sum_{i=1}^n (q_i - 1)\, g_i(\theta)\pi_i^{-1}\right)^\top$, with $q_i = (1-\pi_i)^{1/2}$. In this case, $\widehat{m}_i = \pi_i^{-1}$. Let $\widehat{m}_i^*(\theta)$ be defined by

$$\widehat{m}_i^*(\theta) = \left(\pi_i + \boldsymbol{\eta}^{*\top}\boldsymbol{c}_i^*\right)^{-1} \tag{3.8.1}$$

where $\boldsymbol{\eta}^{*\top}$ is such that $\sum_{i=1}^n \widehat{m}_i^*(\theta)\,\boldsymbol{c}_i^* = \boldsymbol{C}^*$ holds. We propose to use the same empirical log-likelihood ratio function (3.3.1), even though the $\widehat{m}_i^*(\theta)$ do not maximise (3.2.1) exactly, except under equal probability sampling. In other words, Equation (3.8.1) is not a result of the maximisation of (3.2.1). The empirical log-likelihood ratio function is still defined by (3.3.1) with $\ell(\widehat{m}) = \sum_{i=1}^n \log(\widehat{m}_i)$ and $\ell(\widehat{m}^*, \theta) = \sum_{i=1}^n \log(\widehat{m}_i^*(\theta))$.

Corollary 3.8.1. *Let* $\boldsymbol{c}_i = Nn^{-1}\pi_i$, $\boldsymbol{C} = N$, $\boldsymbol{c}_i^* = q_i(Nn^{-1}\pi_i, g_i(\theta))^\top$ *and* $\boldsymbol{C}^* = (\sum_{i=1}^n Nn^{-1}q_i, \sum_{i=1}^n (q_i - 1)\, g_i(\theta)\pi_i^{-1})^\top$. *We have that*

$$\widehat{r}(\theta_0) = \widehat{G}_\pi(\theta_0)^2\widehat{V}[\widehat{G}_\pi(\theta_0)]^{-1} + o_p(1) \tag{3.8.2}$$

where

$$\widehat{V}[\widehat{G}_\pi(\theta_0)] = \sum_{i=1}^n q_i^2\breve{g}_i(\theta_0)^2 - \widehat{d}^{-1}\mathring{G}(\theta_0)^2 \tag{3.8.3}$$

where $\breve{g}_i(\theta_0) = g_i(\theta_0)/\pi_i$ *is the Hájek (1964) variance estimator,* $\mathring{G}(\theta_0) = \sum_{i=1}^{n} q_i^2 \breve{g}_i(\theta_0)$ *and* $\widehat{d} = \sum_{i=1}^{n} q_i^2$.

Proof. Suppose we have a single stratum ($H = 1$), we have that Equation (3.8.2) holds as Equation (3.8.5) of Corollary 3.8.2 reduces to (3.8.3) when $H = 1$. □

If this variance estimator is consistent, we have that $\widehat{r}(\theta)$ follows a chi-squared distribution, by Slutsky's theorem. Hence empirical likelihood confidence intervals can be constructed with $\widehat{r}(\theta)$.

Because we are only interested in the asymptotic behaviour of $\widehat{r}(\theta)$, we only need a consistent variance estimator in (3.8.2). Then a Sen-Yates-Grundy variance estimator (Sen, 1953; Yates and Grundy, 1953) is not necessary. The variance estimator (3.8.3) is a consistent estimator for the variance, for high entropy sampling designs when $d = \sum_{i \in U} \pi_i (1 - \pi_i) \to \infty$ (e.g. Hájek, 1964, 1971; Berger 1998; Deville 1999; Brewer 2002; Brewer and Donadio 2003; Haziza et al. 2004; Tillé 2006; Prášková and Sen 2009; Fuller 2009; Berger 2011). For example, the rejective (Hájek, 1964; Fuller, 2009), the Rao-Sampford (Rao, 1965; Sampford, 1967), the Chao (1982) and the Pareto Aires (2000) sampling designs are high entropy designs (Berger, 2011). Note that most sampling designs used in practice have large entropy, except the non-randomized systematic sampling and the Rao et al. (1962) sampling design. Nevertheless, in Section 3.7, we show that the proposed approach is valid under the Rao et al. (1962) sampling design. For non-randomized systematic sampling, we suggest to use the approach proposed by Berger (2005a) where $\widehat{\boldsymbol{X}_1} = \sum_{i=1}^{n} x_{i1}$ and x_{i1} are values of a variable Υ_1 associated with the i-th unit. By setting $x_{i1} = 1$ then $\widehat{\boldsymbol{X}_1} = n$ and a fixed sample size constraint is kept.

The q_i reduce the effect on the confidence interval of units with large π_i. For example, if $\pi_i = 1$, then $\widehat{m}_i \pi_i = \widehat{m}_i^* (\theta) \pi_i = 1$. This implies that this unit will have no contribution towards the empirical likelihood functions and any confidence intervals. This is a natural property as this unit does not contribute towards the sampling distribution. Note that with small sampling fractions, $q_i \doteq 1$ and when $q_i = 1$, the approach proposed in this section reduces to the approach of Section 3.4. Note that we propose to adjust the constraints by quantities which do not need to be estimated, unlike the pseudo likelihood approach which adjust the empirical log-likelihood ratio function by a quantity that needs to be estimated (the design effect).

For stratified designs, we propose to use $\boldsymbol{c}_i = Nn^{-1}\boldsymbol{z}_i, \boldsymbol{C} = Nn^{-1}\boldsymbol{n}, \boldsymbol{c}_i^* = q_i(Nn^{-1}\boldsymbol{z}_i^\top, g_i(\theta))^\top$ and $\boldsymbol{C}^* = \left(\sum_{i=1}^n Nn^{-1}q_i\breve{\boldsymbol{z}}_i^\top, \sum_{i=1}^n (q_i - 1)\,\breve{g}_i(\theta)\right)^\top$ where $\breve{\boldsymbol{z}}_i = \boldsymbol{z}_i\pi_i^{-1}$. We also consider that $\widehat{m}_i^*(\theta)$ is defined by (3.8.1) and is such that $\sum_{i=1}^n \widehat{m}_i^*(\theta)\,\boldsymbol{c}_i^* = \boldsymbol{C}^*$ holds.

Corollary 3.8.2. *Let* $\ell(\widehat{m}) = \sum_{i=1}^n \log(\widehat{m}_i)$ *with* $\boldsymbol{c}_i = Nn^{-1}\boldsymbol{z}_i$*,* $\boldsymbol{C} = Nn^{-1}\boldsymbol{n}$*. Let* $\ell(\widehat{m}^*, \theta_0) = \sum_{i=1}^n \log(\widehat{m}_i^*(\theta_0))$ *with* $\boldsymbol{c}_i^* = q_i(Nn^{-1}\boldsymbol{z}_i, g_i(\theta_0))^\top$ *and* $\boldsymbol{C}^* = (\sum_{i=1}^n Nn^{-1}q_i\breve{\boldsymbol{z}}_i^\top, \sum_{i=1}^n (q_i - 1)\,\breve{g}_i(\theta_0))^\top$*. We have that*

$$\widehat{r}(\theta_0) = 2\left\{\ell(\widehat{m}) - \ell(\widehat{m}^*, \theta_0)\right\} = \widehat{G}_\pi(\theta_0)^2\widehat{V}_{\text{ST}}[\widehat{G}_\pi(\theta_0)]^{-1} + O_p(n^{-\frac{1}{2}}) \tag{3.8.4}$$

where $\widehat{V}_{\text{ST}}[\widehat{G}_\pi(\theta_0)]$ *where is the Hájek (1964) variance estimator, defined by*

$$\widehat{V}_{\text{ST}}[\widehat{G}_\pi(\theta_0)] = \sum_{h=1}^H \left[\sum_{i\in s_h} q_i^2\breve{g}_i(\theta)^2 - \widehat{d}_h^{-1}\mathring{G}_h(\theta_0)^2\right], \tag{3.8.5}$$

where $\widehat{d}_h^{-1} = \sum_{i\in s_h} q_i^2$ *and* $\mathring{G}_h(\theta_0) = \sum_{i\in s_h} q_i^2\breve{g}_i(\theta)$*.*

Proof. By replacing ψ_i by q_i and $\psi_i^\bullet$ by q_i and using Theorem B.5, we have that $\widehat{\boldsymbol{\Sigma}}_{zz} = N^2n^{-2}\text{diag}[\widehat{d}_1, \ldots, \widehat{d}_h, \ldots, \widehat{d}_H]$, $\widehat{\boldsymbol{\Sigma}}_{zg} = Nn^{-1}[\mathring{G}_1(\theta_0), \ldots, \mathring{G}_h(\theta_0), \ldots, \mathring{G}_H(\theta_0)]^\top$. Thus, $\widehat{\boldsymbol{\Sigma}}_{zg}^\top\widehat{\boldsymbol{\Sigma}}_{zz}^{-1}\widehat{\boldsymbol{\Sigma}}_{zg} = \sum_{h=1}^H \mathring{G}_h(\theta_0)^2\widehat{d}_h^{-1}$. As $\widehat{\sigma}_{gg} = \sum_{h=1}^H \sum_{i\in s_h}(1 - \pi_i)\breve{g}_i(\theta_0)^2$, we have that $\widehat{\sigma}_{gg} - \widehat{\boldsymbol{\Sigma}}_{zg}^\top\widehat{\boldsymbol{\Sigma}}_{zz}^{-1}\widehat{\boldsymbol{\Sigma}}_{zg} = \widehat{V}_{\text{ST}}[\widehat{G}_\pi(\theta_0)]$. Furthermore, $\widehat{m}_i = \pi_i^{-1}$, $-\ell(\widehat{m}) = \sum_{i=1}^n \log(\pi_i) = -\ell(\pi)$ and we have that $\widehat{r}(\theta_0) = -2[\ell(\widehat{m}^*, \theta_0) + \ell(\pi)]$. The corollary is obtained from (B.5.1).

□

This variance estimator is consistent when $d_h = \sum_{i\in U_h} \pi_i\,(1 - \pi_i) \to \infty$ and when the number of strata is bounded. Hence $\widehat{r}(\theta_0)$ follows a chi-squared distribution asymptotically.

The approach described in this section can be extended for calibrations constraints. For $\ell(\widehat{m})$, we propose to use $\boldsymbol{c}_i = q_i(Nn^{-1}\boldsymbol{z}_i^{\mathsf{T}}, \boldsymbol{f}_i(x_i, \boldsymbol{\vartheta}_0)^{\mathsf{T}})^{\mathsf{T}}$ and $\boldsymbol{C} = q_i(\sum_{i=1}^{n} Nn^{-1}\breve{\boldsymbol{z}}_i^{\mathsf{T}}, \sum_{i=1}^{n}(q_i - 1)\breve{\boldsymbol{f}}_i(\boldsymbol{x}_i, \boldsymbol{\vartheta}_0)^{\mathsf{T}})^{\mathsf{T}}$, where $\breve{\boldsymbol{f}}_i(\boldsymbol{x}_i, \boldsymbol{\vartheta}_0) = \boldsymbol{f}_i(\boldsymbol{x}_i, \boldsymbol{\vartheta}_0)\pi_i^{-1}$ and $\boldsymbol{z}_i$ is defined by (3.6.1). Using Theorem (3.7.1) we have that $\widehat{r}(\theta_0)$ follows a chi-squared distribution asymptotically with one degree of freedom.

3.9 Empirical likelihood approach for the Rao-Hartley-Cochran strategy

The Rao-Hartley-Cochran (RHC) sampling design (Rao et al., 1962) is a popular unequal probability sampling and without replacement design which does not belong to the class of high entropy sampling designs. The sampling scheme divides the population N into as many groups as the sample size which provides an unbiased estimate of population totals more efficient, for a fixed sample size, than the estimate based on PPS sampling. Single-stage and two-stage designs are considered by this approach. The procedure provides exact variance and variance estimator expression for any population size N and sample size n. Additionally, the method provides an unbiased estimation of variance, always positive for any sample size. Bansal and Singh (1986) proposed to divide the population N into $(n + k)$ random groups where k is a positive integer. Then a simple random sample without replacement of n groups is selected. The final sample of n units is obtained with probability proportional to p_i from each of the n groups independently. This strategy is more efficient than the usual RHC sampling design if the PPS sampling design is inferior to with replacement sampling with equal probabilities.

Suppose that the population of N units is divided randomly into n groups $A_1, \ldots, A_n$ of sizes $N_1, \ldots, N_n$, with $\sum_{i=1}^{n} N_i = N$. One unit is selected independently from each of the n groups with probability $p_i = \pi_i/a_i$ for the ith group, where $a_i = \sum_{j \in A_i} \pi_j$. The estimator of a total Y is $\widehat{Y} = \sum_{i=1}^{n} y_i/p_i$, where (y_i, p_i) denotes the value for the unit selected form the ith group. The sampling variance of $\widehat{Y}$ under RHC sampling design is given by $V\left[\widehat{Y}\right] = \left[N\left(N\right)\right]^{-1}\left(\sum_{i=1}^{n} N_i^2 - N\right)\left(\sum_{i=1}^{N} y_i^2/\pi_i - Y^2\right)$. This variance is minimum when all the N_i are equal. An unbiased estimator of variance of $\widehat{Y}$ is $\widehat{V}[\widehat{Y}] = \wp^2 \sum_{i=1}^n p_i \left(y_i/\pi_i - \widehat{Y}\right)^2$ where $\wp^2 = \left(\sum_{i=1}^n N_i^2 - N\right) / \left(N^2 - \sum_{i=1}^n N_i^2\right)$. In the case of equal group size ($N_i = N/n$), $\wp$ becomes $(1 - n/N)^{1/2} (n-1)^{-1/2}$. As units are selected independently, the empirical likelihood is given by (3.2.1). We consider that the constraint (3.2.2) is such that the constraint $\sum_{i=1}^n m_i p_i = n$ always hold. By maximising (3.2.1) with $\boldsymbol{c}_i = p_i$ and $\boldsymbol{C} = n$, we obtain that $\widehat{m}_i = p_i^{-1}$. The components of $\boldsymbol{c}_i$ have to be multiplied by Nn^{-1} for the regularity conditions (3.3.6) and (3.3.7) hold. When $\boldsymbol{c}_i = p_i$ condition (3.3.7) does not hold. However, the $\widehat{m}_i$ obtained from $\boldsymbol{c}_i = Nn^{-1}p_i$ and $\boldsymbol{C} = N$ are the same as those obtained when $\boldsymbol{c}_i = p_i$ and $\boldsymbol{C} = n$. When $g_i(\theta) = y_i - n^{-1}p_i\theta$, the maximum empirical likelihood estimator $\widehat{\theta}$, defined by (3.3.2), is the RHC estimator of a total, $\widehat{G}(\theta) = \widehat{G}_{\text{RHC}}(\theta)$, where

$$\widehat{G}_{\text{RHC}}(\theta) = \sum_{i=1}^{n} \frac{g_i(\theta)}{p_i}. \tag{3.9.1}$$

For the computation of confidence intervals, we propose to use $\boldsymbol{c}_i^* = (Nn^{-1}q_i^\circ p_i, q_i^\bullet g_i(\theta))^\top$ and $\boldsymbol{C}^* = \left(\sum_{i=1}^n Nn^{-1}q_i^\circ, \sum_{i=1}^n (q_i^\bullet - 1)\, g_i(\theta)p_i^{-1}\right)^\top$, with $q_i^\circ = a_i^{1/2}$ and $q_i^\bullet = \left(\widehat{\varsigma} n a_i^{-1}\right)^{1/2}$ where $\widehat{\varsigma} = \left(\sum_{i=1}^n N_i^2 - N\right)[N^2 - \sum_{i=1}^n N_i^2]$ is the finite population correction proposed by Rao et al. (1962). We also consider $\widehat{m}_i^*(\theta)$ is defined by (3.8.1) and is such that $\sum_{i=1}^n \widehat{m}_i^*(\theta)\, \boldsymbol{c}_i = \boldsymbol{C}$ holds.

Corollary 3.9.1. *Let* $\boldsymbol{c}_i^* = (Nn^{-1}q_i^\circ p_i, q_i^\bullet g_i(\theta))^\top$ *and* $\boldsymbol{C}^* = (Nn^{-1} \sum_{i=1}^n q_i^\circ, \sum_{i=1}^n (q_i^\bullet - 1)\, g_i(\theta)p_i^{-1})^\top$. *We have*

$$\widehat{r}(\theta_0) = \widehat{G}_{\text{RHC}}(\theta_0)^2 \widehat{V}[\widehat{G}_{\text{RHC}}(\theta_0)]^{-1} + O_p(n^{-1/2}) \tag{3.9.2}$$

and

$$\widehat{V}[\widehat{G}_{\text{RHC}}(\theta_0)] = \widehat{\varsigma}\left(n \sum_{i=1}^{n} a_i \frac{g_i(\theta_0)^2}{\pi_i^2} - \widehat{G}_{\text{RHC}}(\theta_0)^2 \right). \tag{3.9.3}$$

The Rao et al. (1962) estimator $\widehat{G}_{\text{RHC}}(\theta_0)$ is asymptotically normal distributed under the regularity conditions proposed by Ohlsson (1986). Hence $\widehat{r}(\theta_0)$ follows asymptotically a chi-squared distribution with one degree of freedom.

Proof. We have that $\widehat{\boldsymbol{C}}^*_\pi - \boldsymbol{C}^* = (0, \widehat{G}_{\text{RHC}}(\theta_0))^\mathsf{T}$, where $\widehat{\boldsymbol{C}}^*_\pi = \sum_{i=1}^n \boldsymbol{c}_i^* p_i^{-1}$. Using Lemmas B.3 and B.4 with $\psi_i = q_i^\circ$, $\boldsymbol{z}_i = \pi_i$, $\boldsymbol{b}_i = q_i^\bullet(g_i(\theta_0))$ and $\boldsymbol{b}_i^\bullet = (q_i^\bullet - 1)g_i(\theta_0)p_i^{-1}$, we have that $\widehat{r}(\theta_0) = \widehat{G}_{\text{RHC}}(\theta)^2(\widehat{\sigma}_{gg} - \widehat{\boldsymbol{\Sigma}}_{cg}^\mathsf{T}\widehat{\boldsymbol{\Sigma}}_{cc}^{-1}\widehat{\boldsymbol{\Sigma}}_{cg})^{-1} + O_p(n^{-\frac{1}{2}})$ where $\widehat{\boldsymbol{\Sigma}}_{cc} = N^2n^{-2}\sum_{i=1}^n q_i^{\circ 2} = N^2n^{-2}\sum_{i=1}^n a_i = N^2n^{-2}n$, $\widehat{\boldsymbol{\Sigma}}_{cg} = Nn^{-1}\sum_{i=1}^n q_i^\circ q_i^\bullet g_i(\theta_0)p_i^{-1} = Nn^{-1}(\widehat{\varsigma}n)^{\frac{1}{2}}\widehat{G}_{\text{RHC}}(\theta_0)$ and $\widehat{\sigma}_{gg} = \sum_{i=1}^n q_i^\bullet g_i(\theta_0)^2 p_i^{-2} = \widehat{\varsigma}n\sum_{i=1}^n a_i g_i(\theta_0)^2\pi_i^{-2}$. Thus, $\widehat{\sigma}_{gg} - \widehat{\boldsymbol{\Sigma}}_{cg}^\mathsf{T}\widehat{\boldsymbol{\Sigma}}_{cc}^{-1}\widehat{\boldsymbol{\Sigma}}_{cg} = \widehat{V}[\widehat{G}_{\text{RHC}}(\theta_0)]$. Thus, we have that $\widehat{r}(\theta_0) = \widehat{G}_{\text{RHC}}(\theta_0)^2\widehat{V}[\widehat{G}_{\text{RHC}}(\theta_0)]^{-1} + O_p(n^{-1/2})$. □

The approach proposed in this section can be generalised to take in account the stratification and the auxiliary variables (see Theorem 3.7.1).

3.10 Empirical likelihood approach compared with calibration

The calibration approach can be defined as an adjustment method for the original sampling design weights $d_i = \pi_i^{-1}$ by incorporating auxiliary information, with the purpose of reduce the variance of sampling error and increase the precision of estimates. Distances between the original weights and calibrated weights are minimised according to a distance function subject to a set of constraints called calibration equations (Särndal, 2007). When the calibrated weights are applied to the auxiliary variable values, the sample sum of the units agree with the known auxiliary information. This characteristic is appealing because a strong correlation between the auxiliary variables and the variable of interest implies that a good estimation of the auxiliary variables, using the weights, also should estimate well the variable of interest. Formally, a population total $\sum_{i=1}^N \boldsymbol{x}_i = \boldsymbol{T}_x$ for a vector of auxiliary information $\boldsymbol{x}_i$ is known. The calibration estimator of $\overline{Y}$ is $\overline{Y}_{\text{C}} = N^{-1}\sum_{i\in s} w_i y_i$ defines a calibration estimator of $\overline{Y}$ where the calibrated weights w_i minimise the distance function $\Phi_i(w, d)$ between w_i and the basic design weights $d_i = \pi_i$ subject to $\sum_{i\in s} w_i \boldsymbol{x}_i = \boldsymbol{X}$. $\Phi_i(d, w)$ satisfies the following properties: $\Phi_i(d, d) = 0$, it is non-negative, differentiable, strictly convex, with continuos derivative $\phi_i(w,d) = \partial\Phi_i(w,d)/\partial w$ such that $\phi_i(w,d) = 0$ (Deville and Särndal, 1992). Usually the distance function is chosen such that $\phi_i(w,d) = \phi(w,d)/q_i$, where q_i is the predefined positive scale factor, $\phi(\cdot)$ is a function of a single argument, continuos, strictly increasing with $\phi(1) = 0$, $\phi'(1) = 1$ (Särndal, 2007). Let $F^*(z) = \phi^{-1}(z)$ be the inverse function of $\phi(\cdot)$. Minimising $\sum_{i=1}^{n} \Phi_i(w_i, d_i)$ subject to the calibration equation $\sum_{i\in s} w_i \boldsymbol{x}_i = \sum_{i\in U} \boldsymbol{x}_i$ leads to

$$w_i = d_i F^*(q_i, \boldsymbol{x}_i^\top \boldsymbol{\lambda}) \tag{3.10.1}$$

where $\boldsymbol{\lambda}$ is a vector of Lagrange multipliers obtained as the solution of $\sum_{i\in s} d_i \boldsymbol{x}_i F^*(q_i, \boldsymbol{x}_i^\top \boldsymbol{\lambda}) = \sum_{i\in U} \boldsymbol{x}_i$. The chi square distance function $\Phi(w,d) = \sum_{i\in s}(w_i - d_i)^2/d_i q_i$ produces a generalised regression (GREG) estimator of $\overline{Y}$ (Särndal et al., 1992). However, the calibrated weights can be negative under unbalanced sample configurations. The use of other distance functions that force the weights to be positive, involves other disadvantages such as no guarantee of convergence to a solution, very large weights or extreme values of the weights. Furthermore, calibration methods do not produce confidence intervals.

The pseudo empirical likelihood approach resembles a calibration method. The pseudo empirical likelihood function is related to the Kullback-Leibler distance (Rao and Wu, 2009) or minimal entropy distance (Deville and Särndal, 1992) between the pseudo empirical likelihood weights and sampling design weights; the function of the proposed approach (3.2.1) has a likelihood-based motivation and it is not related to a distance function because it does not involve the inclusion probabilities π_i.

The proposed empirical likelihood approach weights $\widehat{m}_i$ are inherently positive. The most complicate computational task is finding the value of the vector $\boldsymbol{\eta}$ presented in Chapter 6 a modified Newton-Raphson algorithm (Polyak, 1987) to estimate the vector $\boldsymbol{\eta}$. The proposed empirical likelihood produces confidences intervals, presented in Chapter 6, with several advantages over the standard confidence intervals depending on variance estimation and bootstrap confidence intervals. Benchmark information and constraints are included in empirical likelihood point and interval estimation. The sampling distribution of data determines the orientation of the empirical likelihood confidence intervals and the range of the parameter space is preserved (Rao and Wu, 2009).

Calibration induces similar asymptotic properties as the empirical likelihood approach. By replacing the empirical log likelihood by any distance functions, results similar to these for empirical likelihood can be obtained. This can be verified using the Euclidean likelihood (Owen, 1991), Kullback-Leibler and Hellinger distances (Owen, 2001, Chapter 3).

Calibration weights (3.10.1) do not necessarily produce an optimal calibration estimator (Kott, 2009) and such optimality property is weak since there are many possible distance functions and the scale factors q_i (Särndal, 2007). The calibration estimator proposed by Rao (1994) and the empirical likelihood estimator are asymptotically optimal (See Section 3.7), given that $\widehat{B}_x$ is the estimator of the covariance between $\widehat{G}_\pi(\theta)$ and $\widehat{X}_\pi$ divided by the estimator of the variance of $\widehat{X}_\pi$ under a with replacement PPS sampling design (e.g. Isaki and Fuller, 1982; Montanari, 1987; Rao, 1994; Särndal, 1996; Berger et al., 2003) when the sampling fraction is small.

There is a analogy between the proposed empirical likelihood and calibration (e.g Huang and Fuller, 1978; Deville and Särndal, 1992), as the function (3.2.1) can be viewed as a calibration distance function, and the empirical likelihood estimator is asymptotically equivalent to a calibrated regression estimator (3.3.11). The distance functions used in calibration are disconnected from the mainstream statistical theory. However, the proposed empirical log-likelihood function (3.2.1) is related to the concept of likelihood. The advantage of the proposed empirical likelihood approach over standard calibration is the fact that (3.2.1) can be used to construct likelihood ratio confidence intervals.

With the proposed empirical likelihood approach $\boldsymbol{X}$ is not necessarily a vector of population total and can be any population parameter. Calibration applies information from known population total. For example the proposed empirical likelihood approach can be used to calibrate to quantiles, ratios or means. This idea comes from (Owen, 2001). There is also a recent application (Lesage, 2011) to use this approach to calibrate on complex parameters.

The asymptotic results derived for calibration estimators by Deville and Särndal (1992) rely on three regularity conditions, which imply that the covariance matrix $nN^{-2}V(\widehat{\boldsymbol{t}}_{x\pi})$ converges to a fixed matrix $\boldsymbol{K}$ and $n^{-1/2}N^{-1}\ \|\widehat{\boldsymbol{t}}_{x\pi}-\boldsymbol{t}\|= O_p(1)$. This condition is the same as $N^{-1}\ \|\widehat{\boldsymbol{C}}_{\pi}-\boldsymbol{C}\|= O_p(n^{-\frac{1}{2}})$ when $\boldsymbol{c}_i = \boldsymbol{x}_i$. That comes from (3.3.3), one of regularity conditions of the proposed empirical likelihood approach. They also assumed that $n^{-1/2}N^{-1}(\widehat{\boldsymbol{t}}_{x\pi} - \boldsymbol{t})$ converges to the multivariate normal $N(\boldsymbol{0}, \boldsymbol{K})$. For empirical likelihood this condition is not needed. However, for confidence intervals we need a regularity condition similar as the condition (3.4.1).

3.11 Simulation Studies

We generated several population data according to the following model proposed by Wu and Rao (2006)

$$y_i = 3 + a_i + \beta x_i + \varphi e_i, \tag{3.11.1}$$

where a_i and x_i follow independent exponential distributions with rate parameters equal to one and $e_i \sim \chi^2_1 - 1$. The π_i are proportional to $a_i + 2$. The constant 2 is added to a_i to avoid having very small π_i. Populations of size $N = 800$ and $N = 150$ were generated using (3.11.1). The values y_i, x_i and a_i generated were treated as fixed. The parameter φ was used to obtain a weak and a strong correlation between the values y_i and $\hat{y}_i = 3 + a_i + \beta x_i$. Let $\rho\,(y_i, \hat{y}_i)$ denote the correlation. The parameter β was equal to one or zero. The parameter of interest θ_0 is the population mean. For the proposed approach we used $g_i(\theta) = y_i - n^{-1}\theta\pi_i$. Note the pseudo empirical likelihood point estimators EL1 and EL2 are different from the proposed estimators.

We used Chao (1982) sampling design to select 1000 samples with unequal probabilities in order to compare the performance of the 95% empirical likelihood confidence interval with the standard confidence interval based on the central limit theorem, the pseudo empirical likelihood (EL1 and EL2) confidence interval proposed by Wu and Rao (2006), the rescaled bootstrap (Rao and Wu, 1988; Rao et al., 1992) and direct bootstrap (Antal and Tillé, 2011). We consider that the population has a single stratum.

The Chao (1982) sampling scheme is free of second-order inclusion probabilities and it only depends on the first-order inclusion probabilities of the sampled units (Berger, 2005b). This property is particularly useful for the variance estimator involved in the construction of the standard and pseudo empirical likelihood confidence intervals. The Sen-Yates-Grundy variance estimator (Sen, 1953; Yates and Grundy, 1953) is used for the standard confidence intervals and the pseudo empirical likelihood approach. The simulations were programmed in the statistical software R (R Core Team, 2012). Descriptions of main function are included in Chapter 6 and Appendix A. The observed coverages, the lower and upper tail error rates (E.R.) and the average length of the 95% confidence intervals and the relative mean square error (Relative mean square error (RRMSE)) of the point estimator are reported in Tables 3.1 and 3.2.

We considered two point estimators. In the first case, we defined $\boldsymbol{c}_i = \pi_i$ and for the second case we used $\boldsymbol{c}_i = (1, \pi_i)^\top$.We considered the situation when $N = 800$: the coverages and error tail rates are reported in Table 3.1 (values not in brackets). We consider sampling fractions are negligible when $nN^{-1} \leq 10\%$ (see Cochran, 1977, p. 24). With a small sample size ($n = 40$) and a weak correlation ($\rho\left(y_i, \hat{y}_i\right) = 0.30$), we do not observe significant differences between the different approaches compared. In all the cases, the proposed estimator with $\boldsymbol{c}_i = \pi_i$ has a good coverage. When $\rho\left(y_i, \hat{y}_i\right) = 0.80$, the proposed estimator with $\boldsymbol{c}_i = (1, \pi_i)^\top$ is as accurate as the EL2 estimator, and the proposed estimator has a better coverage when the sample size is large. Figure 3.1 shows the Quantile-Quantile plot of the empirical log-likelihood ratio function is well approximated by a chi-squared distribution. We observe a small departure which can be due to the fact the finite population corrections are ignored. However, the coverages of the confidence intervals are not seriously affected because the departures happen mostly after the 0.05 quantile of the χ_1^2 distribution.

In the Table 3.1, we also have results with $n = 150$ (in brackets). In this case the sampling fraction is not negligible as $nN^{-1} > 0.25$. Note that in this case, the standard confidence intervals may have poor coverages and unbalanced tail E.R. For example, when $\rho\left(y_i, \hat{y}_i\right) = 0.30$ and $n = 40$ the standard (Normal) confidence intervals show a poor coverage of 89.6% and most of their tail E.R. are situated in the upper side (10%).

Although the rescaled bootstrap confidence intervals have better coverages than the rest of the approaches, the confidence interval lengths are the largest of the seven methods. When $\rho\left(y_i, \hat{y}_i\right) = 0.80$, the proposed estimator with $\boldsymbol{c}_i = \left(1, \pi_i\right)^{\mathsf{T}}$ is as accurate as the EL2 estimator and the proposed estimator has better coverages when the sample size is large. Note that the proposed estimator with $\boldsymbol{c}_i = \pi_i$ is more accurate that the EL1 estimator. The Quantile-Quantile plots in Figures 3.1 and 3.2, show that the empirical log-likelihood ratio function has a chi-squared distribution when $\boldsymbol{c}_i = \pi_i$, except in cases where correlation $\rho\left(y_i, \hat{y}_i\right) = 0.30$ and $n = 40$ for both population sizes. This concords with the low coverage of the corresponding confidence intervals, 91.5% and 91.4% respectively.

Table 3.1: Coverages of the 95% confidence intervals for means. The x_i are not generated ($\beta = 0$). The values not in brackets are for populations of size $N = 800$ (small sampling fractions). The values in brackets are for populations of size $N = 150$ (large sampling fractions).

$\rho(y_i, \hat{y}_i)$	n	Approaches	Coverage Prob. (%)	Lower Tail E.R. (%)	Upper Tail E.R. (%)	Average Length	RRMSE (%)
0.3	40	Proposed $\boldsymbol{c}_i = \pi_i$	91.5 (91.4)	2.2 (2.3)	6.3 (6.3)	1.96 (1.95)	9.5 (10.1)
		Proposed $\boldsymbol{c}_i = (1, \pi_i)^\top$	91.2 (90.7)	2.5 (2.4)	6.3 (6.9)	1.86 (1.90)	9.6 (10.0)
		Pseudo-EL1	92.9 (91.6)	2.3 (2.0)	4.8 (6.4)	1.97 (1.94)	9.6 (10.0)
		Pseudo-EL2	91.1 (90.6)	2.6 (2.1)	6.3 (7.3)	1.85 (1.88)	9.5 (10.0)
		$N^{-1}\widehat{Y}_\pi$ (Normal)	90.7 (89.6)	0.7 (0.4)	8.6 (10.0)	1.87 (1.90)	9.5 (10.1)
		Rescaled bootstrap	91.1 (93.6)	1.1 (0.3)	7.8 (6.1)	1.90 (2.17)	9.5 (10.1)
		Direct bootstrap	90.0 (89.6)	1.0 (0.5)	9.0 (9.9)	1.86 (1.89)	9.5 (10.1)
	80	Proposed $\boldsymbol{c}_i = \pi_i$	95.0 (92.8)	2.5 (2.9)	2.5 (4.3)	1.42 (1.19)	6.7 (5.8)
		Proposed $\boldsymbol{c}_i = (1, \pi_i)^\top$	92.8 (92.9)	3.5 (3.1)	3.7 (4.1)	1.33 (1.17)	6.7 (5.8)
		Pseudo-EL1	93.5 (93.2)	3.0 (2.0)	3.5 (4.8)	1.38 (1.16)	7.0 (5.8)
		Pseudo-EL2	92.7 (93.0)	3.3 (2.0)	4.0 (5.0)	1.31 (1.15)	6.7 (5.8)
		$N^{-1}\widehat{Y}_\pi$ (Normal)	92.3 (93.5)	1.5 (0.8)	6.2 (5.7)	1.33 (1.15)	6.7 (5.8)
		Rescaled bootstrap	94.2 (97.6)	1.5 (0.0)	4.3 (2.4)	1.39 (1.59)	6.7 (5.8)
		Direct bootstrap	92.1 (93.4)	1.5 (0.8)	6.4 (5.8)	1.33 (1.15)	6.7 (5.8)
0.8	40	Proposed $\boldsymbol{c}_i = \pi_i$	94.2 (95.0)	2.1 (1.7)	3.7 (3.3)	0.62 (0.48)	3.0 (2.4)
		Proposed $\boldsymbol{c}_i = (1, \pi_i)^\top$	91.1 (94.8)	2.8 (1.6)	6.1 (3.6)	0.47 (0.39)	2.4 (1.9)
		Pseudo-EL1	94.7 (95.4)	2.6 (2.0)	2.7 (2.6)	0.76 (0.62)	3.8 (3.0)
		Pseudo-EL2	91.8 (93.9)	2.9 (1.8)	5.3 (4.3)	0.45 (0.38)	2.4 (1.9)
		$N^{-1}\widehat{Y}_\pi$ (Normal)	93.3 (94.7)	1.3 (1.1)	5.4 (4.2)	0.59 (0.48)	3.0 (2.4)
		Rescaled bootstrap	94.0 (96.6)	1.5 (0.8)	4.5 (2.6)	0.61 (0.56)	3.0 (2.4)
		Direct bootstrap	92.8 (94.6)	1.5 (1.1)	5.7 (4.3)	0.59 (0.48)	3.0 (2.4)
	80	Proposed $\boldsymbol{c}_i = \pi_i$	95.6 (94.8)	1.4 (2.7)	3.0 (2.5)	0.45 (0.26)	2.0 (1.3)
		Proposed $\boldsymbol{c}_i = (1, \pi_i)^\top$	93.6 (94.6)	2.6 (2.7)	3.8 (2.7)	0.34 (0.22)	1.7 (1.1)
		Pseudo-EL1	95.7 (94.9)	1.7 (2.3)	2.6 (2.8)	0.51 (0.31)	2.5 (1.6)
		Pseudo-EL2	92.4 (93.7)	2.9 (2.4)	4.7 (3.9)	0.32 (0.22)	1.7 (1.1)
		$N^{-1}\widehat{Y}_\pi$ (Normal)	93.9 (93.5)	1.1 (1.4)	5.0 (5.1)	0.41 (0.25)	2.0 (1.3)
		Rescaled bootstrap	94.6 (99.5)	1.1 (0.0)	4.3 (0.5)	0.44 (0.40)	2.0 (1.3)
		Direct bootstrap	93.9 (93.8)	1.1 (1.4)	5.0 (4.8)	0.42 (0.26)	2.0 (1.3)

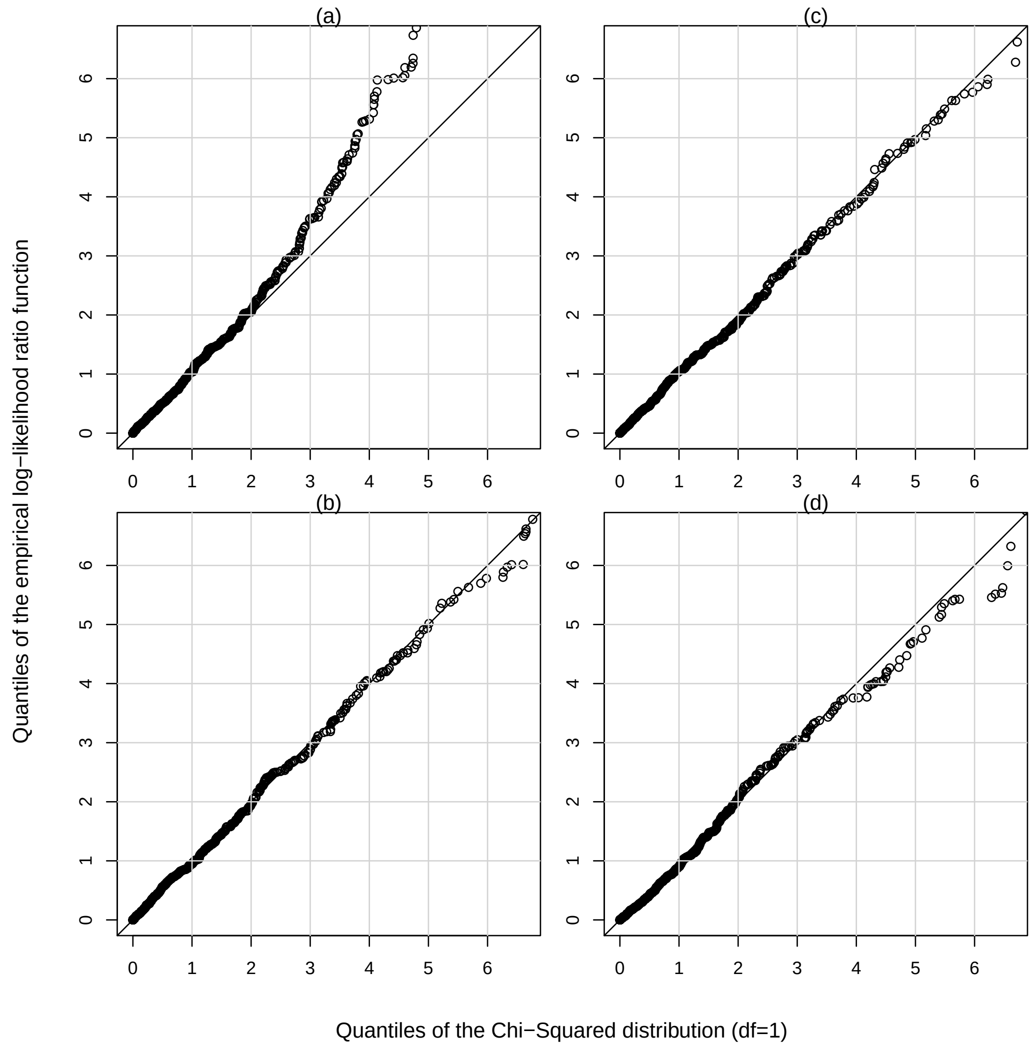

Figure 3.1: Quantile-Quantile plot of the observed distribution of the proposed empirical log-likelihood ratio function $\widehat{r}(\theta_0)$ when $\boldsymbol{c}_i = \pi_i,\ N = 800$; $n = 40$ with (a) $\rho\left(y_i, \hat{y}_i\right) = 0.30$, (b) $\rho\left(y_i, \hat{y}_i\right) = 0.80$; $n = 80$ with (c) $\rho\left(y_i, \hat{y}_i\right) = 0.30$ (d) $\rho\left(y_i, \hat{y}_i\right) = 0.80$. The parameter of interest θ_0 is the population mean. We considered that we have a negligible sampling fraction. The approach of Section 3.4 is used. The data are generated with the model (3.11.1) with $\beta = 0$.

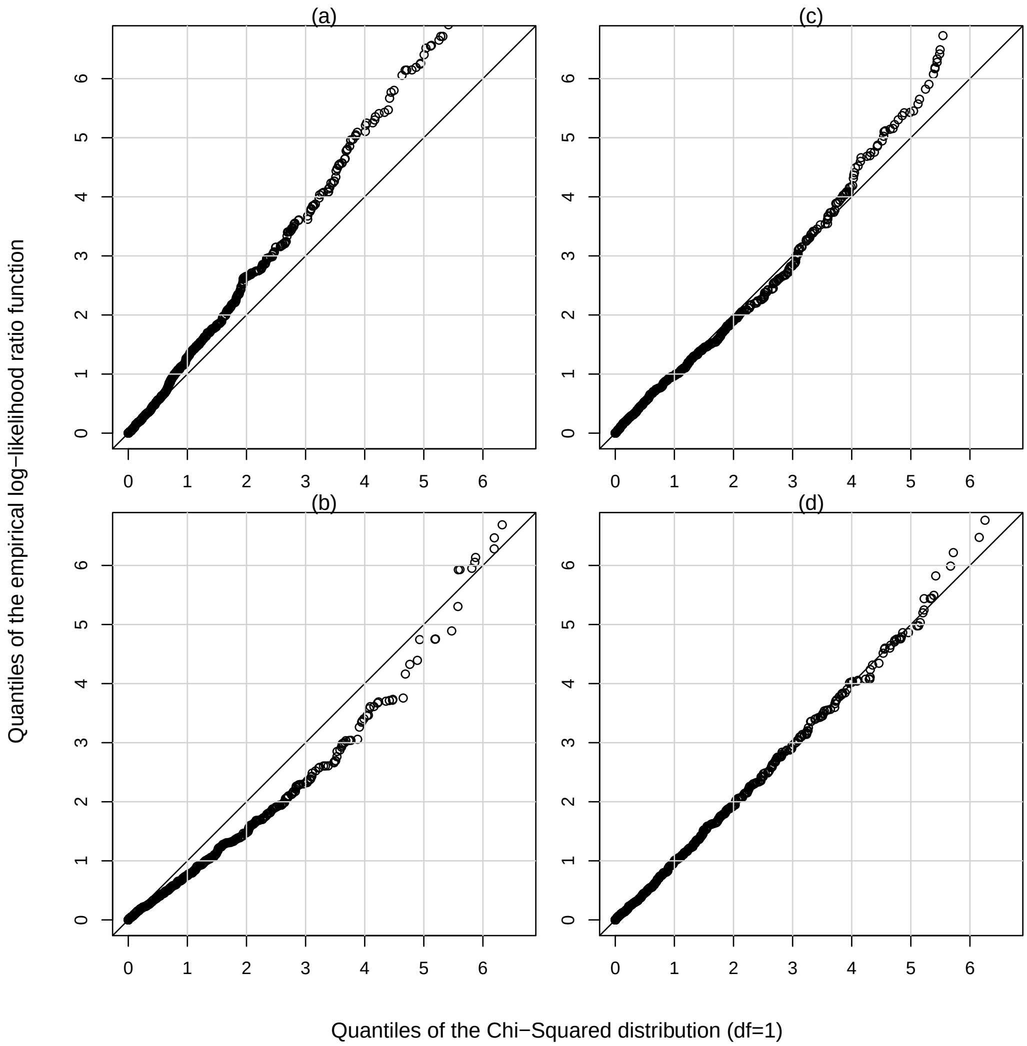

Figure 3.2: Quantile-Quantile plot of the observed distribution of the proposed empirical log-likelihood ratio function $\widehat{r}(\theta_0)$ when $\boldsymbol{c}_i = \pi_i,\ N = 150$; $n = 40$ with (a) $\rho(y_i, \hat{y}_i) = 0.30$, (b) $\rho(y_i, \hat{y}_i) = 0.80$; $n = 80$ with (c) $\rho(y_i, \hat{y}_i) = 0.30$ (d) $\rho(y_i, \hat{y}_i) = 0.80$. The parameter of interest θ_0 is the population mean. We considered that we have a non-negligible sampling fraction. The approach of Section 3.5 is used. The data are generated with the model (3.11.1) with $\beta = 0$.

For the next series of simulations, we generate an auxiliary variable $(\beta = 1)$, and a random set of 80% of the values of y_i are replaced by zero. We consider two point estimators. For the first case we use $\boldsymbol{c}_i = (x_i, \pi_i)^{\mathsf{T}}$, for the second case, we use $\boldsymbol{c}_i = (1, x_i, \pi_i)^{\mathsf{T}}$. The standard confidence interval is based on the standard regression estimator defined $(6.4.2)$ in Särndal et al. (1992). The results are shown in Table 3.2.

The proposed estimator with $\boldsymbol{c}_i = (x_i, \pi_i)^{\mathsf{T}}$ is one of the most accurate estimator. This is due to the optimality of the proposed estimator (see the end of Section 3.3). When the correlation is strong $(\rho\,(y_i, \hat{y}_i) = 0.63)$, the proposed approaches give better coverages than the EL2 approach. The low coverage of the EL2 estimator is probably due to the instability of the linearised variance estimator used in the design effect. This instability is caused by the skewness of the residuals in (3.11.1). The proposed approach gives better coverages because of the units with larger π_i have a negligible effect on the confidence intervals. The overall coverage of the EL1 approach is also satisfactory, although this approach can give less accurate point estimators. The Quantile-Quantile plots, in Figure 3.3 and Figure 3.4, shows that the restricted empirical log-likelihood ratio function follows a chi-squared distribution when $\boldsymbol{c}_i = (\boldsymbol{x}_i, \pi_i)^{\mathsf{T}}$ and $\boldsymbol{c}_i = (1, \boldsymbol{x}_i, \pi_i)^{\mathsf{T}}$, respectively; in both approaches plots in (a) when $n = 40$ and $\rho\,(y_i, \hat{y}_i) = 0.30$, departures from a chi-square distribution are observed. It is also reflected in confidence interval total rate errors evidently below the intended nominal level of 95%.

Table 3.2: Coverages of the 95% confidence intervals for means. $N = 150$. A random set of 80% of the values of y_i are replaced by zero. The x_i are generated $(\beta = 1)$. θ_0 is the population mean.

$\rho(y_i, \hat{y}_i)$	n	Approaches	Coverage Prob. (%)	Lower Tail E. R. (%)	Upper Tail E. R. (%)	(%)
0.30	40	Proposed $\boldsymbol{c}_i = (x_i, \pi_i)^{\mathsf{T}}$	92.5	2.8	4.7	26.5
		Proposed $\boldsymbol{c}_i = (1, x_i, \pi_i)^{\mathsf{T}}$	91.9	2.9	5.2	26.8
		Pseudo-EL1	92.3	2.5	5.2	26.9
		Pseudo-EL2	90.6	2.3	7.1	26.8
		Regression Estimator (Normal)	90.2	1.3	8.5	26.5
	80	Proposed $\boldsymbol{c}_i = (x_i, \pi_i)^{\mathsf{T}}$	93.9	3.2	2.9	14.2
		Proposed $\boldsymbol{c}_i = (1, x_i, \pi_i)^{\mathsf{T}}$	93.4	3.5	3.1	14.4
		Pseudo-EL1	93.6	2.2	4.2	14.4
		Pseudo-EL2	93.1	2.1	4.8	14.4
		Regression Estimator (Normal)	93.4	1.4	5.2	14.3
0.63	40	Proposed $\boldsymbol{c}_i = (x_i, \pi_i)^{\mathsf{T}}$	95.5	1.5	3.0	22.1
		Proposed $\boldsymbol{c}_i = (1, x_i, \pi_i)^{\mathsf{T}}$	93.7	2.0	4.3	22.3
		Pseudo-EL1	94.5	2.3	3.2	23.3
		Pseudo-EL2	86.9	3.7	9.4	22.4
		Regression Estimator (Normal)	91.5	1.3	7.2	22.4
	80	Proposed $\boldsymbol{c}_i = (x_i, \pi_i)^{\mathsf{T}}$	95.3	1.6	3.1	11.8
		Proposed $\boldsymbol{c}_i = (1, x_i, \pi_i)^{\mathsf{T}}$	94.8	1.6	3.6	12.0
		Pseudo-EL1	94.5	1.3	4.2	11.7
		Pseudo-EL2	89.8	1.6	8.6	12.1
		Regression Estimator (Normal)	93.2	1.0	5.8	12.1

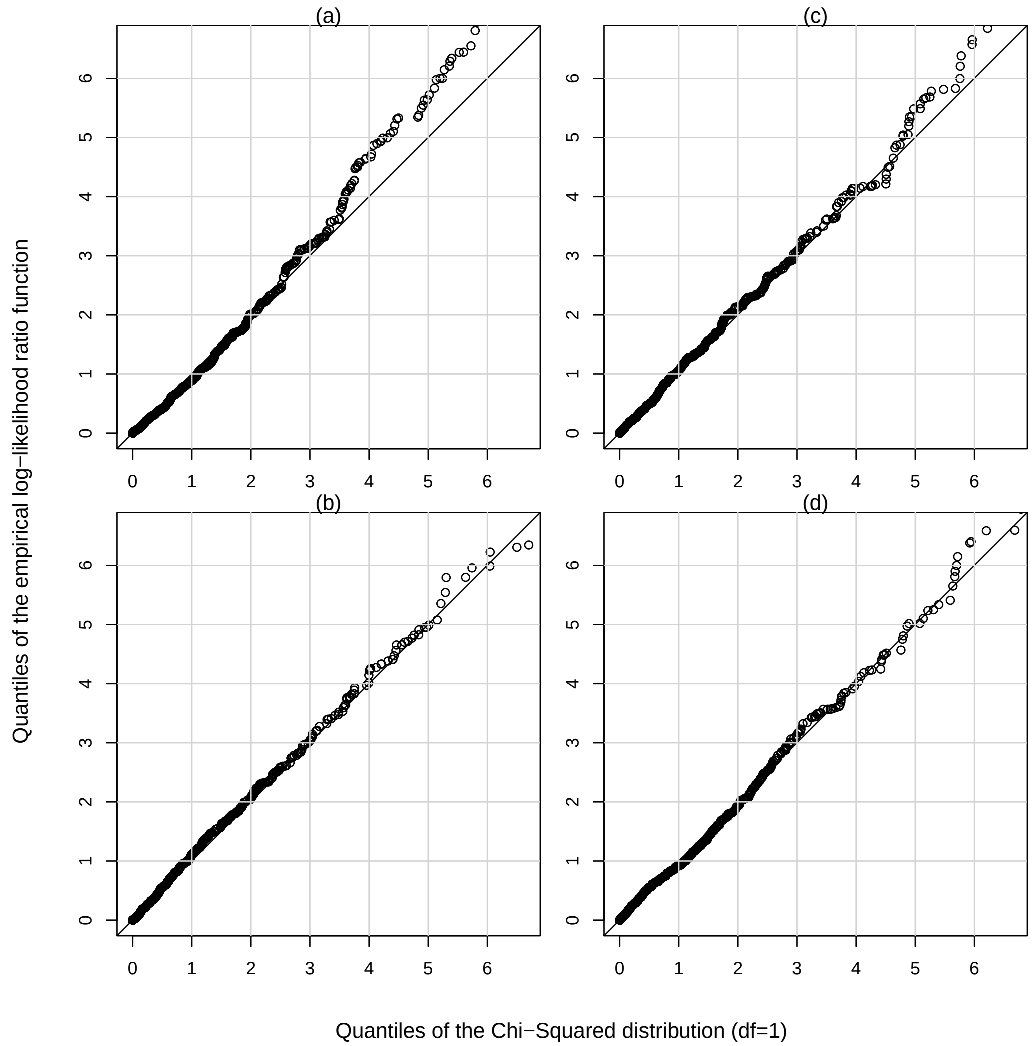

Figure 3.3: Quantile-Quantile plot of the observed distribution of the proposed empirical log-likelihood ratio function $\widehat{r}(\theta_0)$ when $\boldsymbol{c}_i = (x_i, \pi_i)^{\mathsf{T}}$, $N = 150$; $n = 40$ with (a) $\rho(y_i, \hat{y}_i) = 0.30$, (b) $\rho(y_i, \hat{y}_i) = 0.63$; $n = 80$ with (c) $\rho(y_i, \hat{y}_i) = 0.30$ (d) $\rho(y_i, \hat{y}_i) = 0.63$. The parameter of interest θ_0 is the population mean. We considered that we have a non-negligible sampling fraction. The data are generated with the model (3.11.1) with $\beta = 1$. A random set of 80% of values of y_i are replaced by zero.

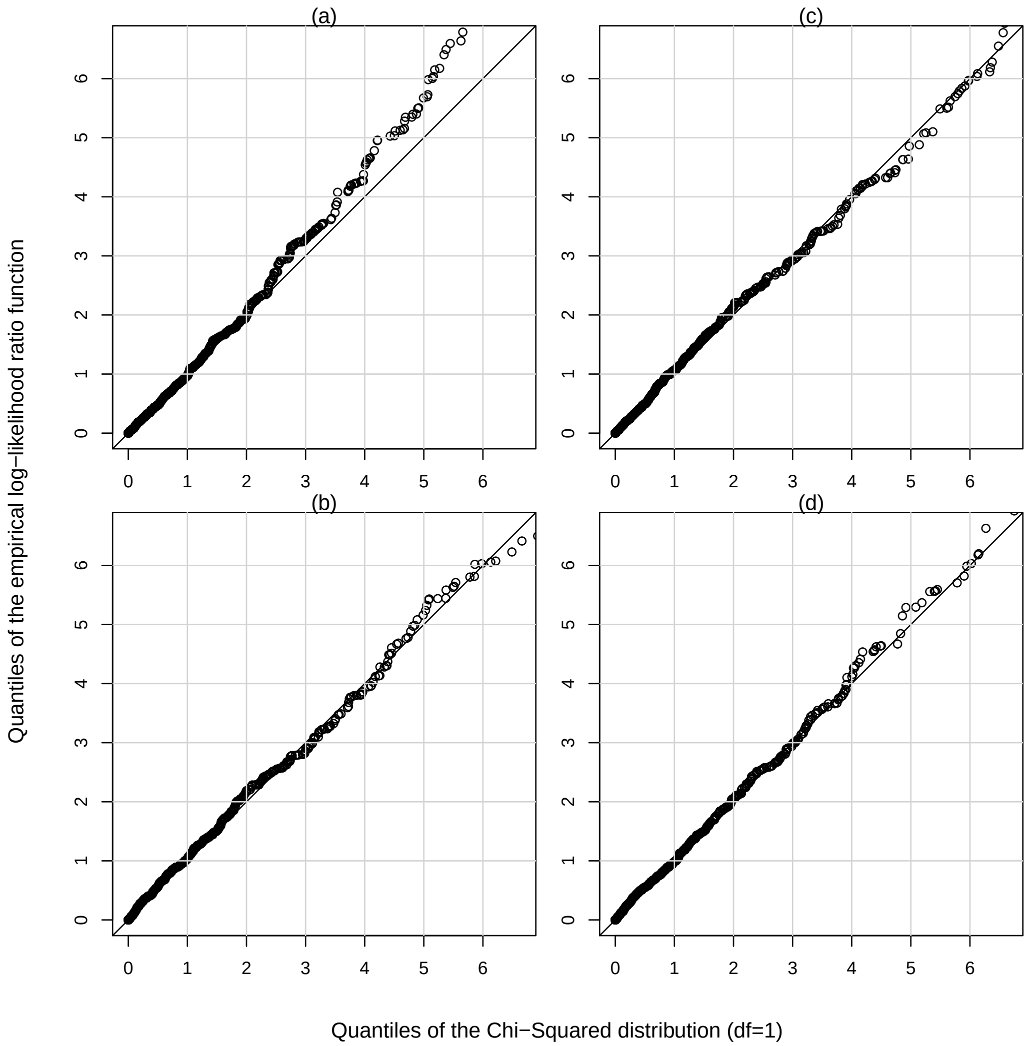

Figure 3.4: Quantile-Quantile plot of the observed distribution of the proposed empirical log-likelihood ratio function $\widehat{r}(\theta_0)$ when $\boldsymbol{c}_i = (1, x_i, \pi_i)^{\mathsf{T}}$, $N = 150$; $n = 40$ with (a) $\rho\,(y_i, \hat{y}_i) = 0.30$, (b) $\rho\,(y_i, \hat{y}_i) = 0.63$; $n = 80$ with (c) $\rho\,(y_i, \hat{y}_i) = 0.30$ (d) $\rho\,(y_i, \hat{y}_i) = 0.63$. The parameter of interest θ_0 is the population mean. We considered that we have a non-negligible sampling fraction. The data are generated with the model (3.11.1) with $\beta = 1$. A random set of 80% of values of y_i are replaced by zero.

3.12 Concluding remarks

Standard confidence intervals based on the central limit theorem and pseudo empirical likelihood confidence intervals require variance estimates which often involve linearisation or resampling. The coverage of standard confidence intervals can be poor with skewed variables. Even if the parameter of interest is not linear, the proposed method does not rely on normality of the point estimator, variance estimates, linearisation, resampling, and joint inclusion probabilities. Empirical likelihood confidence intervals can be easier to compute than standard confidence intervals based on variance estimates. The coverage of the proposed approach is usually better. There is an analogy between the proposed empirical likelihood approach and calibration (Huang and Fuller, 1978; Deville and Särndal, 1992), as the function (3.2.1) can be viewed as a calibration objective function, and the empirical likelihood estimator is asymptotically equivalent to a calibrated regression estimator (3.3.12). The objective functions used for calibration are disconnected from mainstream statistical theory. However, the proposed objective function (3.2.1) is related to the concept of likelihood. The advantage of the proposed empirical likelihood approach over standard calibration is the fact that (3.2.1) can be used to construct likelihood ratio confidence intervals. Empirical likelihood approaches are more general than calibration, and can be used for a wider class of parameters. The bootstrap is an alternative approach which can be used to derive non-parametric confidence intervals. The accuracy of the bootstrap confidence intervals has only been shown theoretically in few particular cases (Rao and Wu, 1988). The proposed approach is simpler to implement and less computationally intensive than the bootstrap, especially with calibration weights. Our simulations study also shows that bootstrap confidence intervals may not have the right coverage and may be more unstable. From a practical point of view, the bootstrap is usually preferred because it does not rely on analytic derivation. The proposed approach possesses the same property, as it does not rely any analytic derivation. Unlike the pseudo empirical likelihood approach, the proposed approach does not rely on variance estimates which could be difficult to estimate for complex parameters. This means that the proposed approach can be applied to a wider class of parameters. The proposed approach is also simpler to implement than the pseudo empirical likelihood. The simulations show that the proposed approach may give better confidence intervals. In cases with small sample size and weak correlation, the Quantile-Quantile plots showed that the empirical log-likelihood ratio function was not properly approximated by a chi-squared distribution. As a consequence, the coverage probability were not close to the nominal coverage. This condition may be amended if the finite population correction is considered.

4

Estimation of empirical likelihood confidence intervals for quantiles

4.1 Introduction

The identification of subgroups lying above or below of the population median for a specific variable is often a goal in survey sampling research. Estimation of the population distribution function and associated quantiles are used to achieve this task. Quantile estimation cannot be considered as a simple function of means, although, relevant information could be obtained by using the relationship between the distribution function and quantiles (Fuller, 2009). Two methods have been used for the estimation of quantiles: inversion of the estimated distribution function and the direct estimation. In both approaches there exist methods to measure the precision of the estimates through the construction of confidence intervals which imply variance estimations. Methods for confidence intervals and variance estimates for distribution functions are grouped in the next three approaches: (i) plug-in model based, estimating asymptotic variance, (ii) design-based and (iii) replication methods (Dorfman, 2009). In the case of quantiles, Woodruff (1952) proposed the first method to build confidence intervals for the median in complex survey designs. The data are assumed to follow a normal distribution. The method can be extended for the other quantiles using a transformation of the cumulative distribution function estimates. The sample quantiles can be expressed asymptotically as linear function of the empirical distribution function evaluated at the quantile q making use of the Bahadur (1966) representation $\theta - \theta_0 = (f(\theta_0))^{-1}\left[q - \widehat{F}(\theta_0)\right] + o_p(n^{-1/2})$, where the parameter θ_0 of interest is the q quantile Y_q of the population distribution of a variable of interest y_i, with $0 < q < 1$; $f(\cdot)$ is the density function of the limiting distribution function of $F(y)$ as $N \to \infty$, Francisco and Fuller (1991) gave sufficient conditions for this representation to hold. For smooth functions of population totals it is possible to apply linearisation techniques for variance estimation such as Taylor linearisation or jackknife linearisation, although for estimation of variance of nonlinear statistics such as quantiles these methods could perform poorly, particularly for small sample sizes. Deville (1999) proposed the generalised linearisation technique for estimating the variance of nonlinear statistics. The basic idea is to approximate certain estimator and nonlinear statistics using Horvitz and Thompson (1952) statistics with the form $\sum_{i=1}^{n} z_i \pi_i^{-1}$ for chosen variables. Hence the variance of linearised variables z_i can be estimated by suitable survey sampling techniques. The method is based on the concept of an influence function (Hampel et al., 1986). The linearised variables of quantiles require to estimate the value of a density function $f(y)$ corresponding to a distribution function $F(y)$. As $f(y)$ assumes a density at the quantile of interest which does not exist because the staircase-shape of $F(y)$, Deville (1999) proposed to approximate the Y_q through Gaussian kernel density estimation

$$\widehat{f}(y) = \frac{1}{\widehat{N}h} \sum_{i=1}^{n} w_i K\left[\frac{(y - y_i)}{h}\right] \tag{4.1.1}$$

where $\widehat{N}$ is an estimation of the population size, the sample weight w_i is the inverse of the inclusion probability π_i of i and h is a band width or average width of the support $K[x] = (2\pi)^{-1/2}\exp(-x^2/2)$. Choices for h were suggested by Silverman (1986) as $h = 0.79(\widehat{Y}_{0.75} - \widehat{Y}_{0.25})\widehat{N}^{-1/5}$ and by Osier (2009) as $h = \widehat{\sigma}\widehat{N}^{-1/5}$ where $\widehat{\sigma}$ is the sample standard deviation of the variable of interest Y. The resulting linearised variable is defined as

$$\widehat{z}_i = -\frac{1}{f(\widehat{Y}_q)}\frac{1}{\widehat{N}}[\delta(y_i \leqslant \widehat{Y}_q) - q]. \tag{4.1.2}$$

Note that $f(\cdot)$ would need to be estimated using a Kernel approach. This implies that $\widehat{z}_i$ will depend on how $f(\cdot)$ is estimated. The variance of the estimator $\widehat{Y}_q$ is approximate by

$$V\left(\sum_{i=1}^{n} \widehat{z}_i w_i\right) \approx V(\widehat{Y}_q). \tag{4.1.3}$$

Chen and Hall (1993) proposed an empirical likelihood confidence intervals for quantiles with identically and independent distributed (iid) observations that require a smoothed version of a degenerate distribution function G_h and an rth-order kernel function K. This approach cannot be implemented directly when the units are selected with unequal probabilities. Under Owen (1988)'s empirical likelihood the confidence interval for a population quantile obtained is the binomial-method interval with the iid observations. The weakness of the binomial-method interval is that its coverage accuracy cannot be higher than $n^{-1/2}$ (Chen and Hall, 1993).

4.2 Estimation of quantiles

As the estimating equation $\sum_{i=1}^{n} \widehat{m}_i \left(\delta\left\{y_i \leq \theta\right\} - q\right) = 0$ where $\delta\left\{y_i \leq \theta\right\} = 1$ if $y_i \leq \theta$ and is equal to zero otherwise, it does not always have a solution and cannot be used to derive an empirical log-likelihood ratio function (e.g. Owen, 2001, p. 45). In order to solve this problem, we proposed to use the following function $g_i(\theta) = \varrho\left(y_{(i)}, \theta\right) - q$, where

$$\varrho\left(y_{(i)}, \theta\right) = \delta\left\{y_{(i)} \leq \theta\right\} + \frac{\theta - y_{(i-1)}}{y_{(i)} - y_{(i-1)}}\delta\left\{y_{(i-1)} \leq \theta\right\}\left(1 - \delta\left\{y_{(i)} \leq \theta\right\}\right) \tag{4.2.1}$$

where $y_{(i)}$ is the value of the i-th sampled units arranged in increasing order, with $y_{(0)} = y_{(1)} - \left(y_{(2)} - y_{(1)}\right)$. The empirical likelihood estimator of Y_q is the solution of the equation $\widehat{G}(\theta) = 0$ which becomes $\tilde{F}(\theta) = q$, where $\tilde{F}(\theta) = \left(\sum_{i=1}^{n} \widehat{m}_{(i)}\right)^{-1} \sum_{i=1}^{n} \widehat{m}_{(i)}\varrho\left(y_{(i)}, \theta\right)$. Note that $\tilde{F}(\theta) = q$ has always a solution because $\tilde{F}(y)$ is a bijective function given by a piecewise linear interpolation of the step distribution function $\widehat{F}(\theta) = \left(\sum_{i=1}^{n} \widehat{m}_{(i)}\right)^{-1} \sum_{i=1}^{n} \widehat{m}_{(i)} \delta\left\{y_{(i)} \leq \theta\right\}$. This interpolation consist of joining the steps of $\widehat{F}(\theta)$ by straight line segments. It can be shown that

$$\frac{\widehat{G}_{\pi}(\theta_0)}{N} = \frac{1}{N}\sum_{i=1}^{n}\frac{1}{\pi_i}\left[\varrho\left(y_{(i)} \leq \theta_0\right) - q\right] \doteq \frac{1}{N}\sum_{i=1}^{n}\frac{1}{\pi_i}\left[\delta\left\{y_i \leq \theta_0\right\} - q\right]$$

which is a Horvitz and Thompson (1952) estimator. Thus, (3.4.1) and the empirical log-likelihood ratio function has asymptotically a chi-square distribution. Therefore, the empirical log-likelihood ratio function can be used to derive confidence intervals for Y_q. In the presence of tied values in the sample the Function (4.2.1) has a undetermined form. In the following paragraph a procedure in presented to adjust observations in the presence of tied data.

We adapted a technique suggested by Owen (2001) for handling ties in the data only to find the solution of (4.2.1). This modification does not affect the original empirical likelihood function. We add to each $y_{(i)}$ the minimum distance between ordered observation $y_{(i)}$ and $y_{(i+1)}$ times a very small quantity, for instance a random uniform number between $-A$ and A where A could be 10^{-6}. The value of $\widehat{\theta}$ can be found using the transformed data and Function (4.2.1). In Section 4.3, we show via simulation that the empirical likelihood confidence intervals have good coverage (see Tables 4.1, 4.2 and 4.3). The approach proposed in this section can be generalised to take into account the stratification and the auxiliary variables (see Sections 3.7 and 3.8).

4.3 Simulation study

We used the same artificial population of Section 3.11 with two values for the correlation: 0.30 and 0.80, and $\rho(y_i, \hat{y}_i)$ denotes the correlation between y_i and $\hat{y}_i$. The parameters of interest θ_0 are the population quantiles Y_q when $q = 0.10$, 0.25 and 0.50. We used the Chao (1982) sampling design to generate 1000 independent samples of size $n = 40$ and $n = 80$. We consider 95% confidence intervals. For the proposed approach, we use $\boldsymbol{c}_i = \pi_i$ and $\boldsymbol{C} = n$; therefore $m_i = \pi_i^{-1}$.

The point estimator is the solution of (3.3.12) with $g_i(\theta) = \varrho\left(y_{(i)}, \theta\right) - q$. This estimator has a skewed sampling distribution. The performance of the proposed empirical likelihood confidence intervals are compared with the direct bootstrap approach (Antal and Tillé, 2011) based on the bootstrap variance, the Woodruff (1952) approach, the rescaled bootstrap approach (Rao and Wu, 1988; Rao et al., 1992) based on the observed confidence intervals based on the percentile method and the standard approach based on linearisation (Deville, 1999). We used 1000 bootstrap replicates in both resampling approaches. For the Woodruff approach, the confidence interval was obtained from the inverse of $\widehat{F}(y) = \widehat{N}_\pi^{-1} \sum_{i=1}^n \pi_i^{-1} \delta \left\{y_i \leq y\right\}$, where $\widehat{N}_\pi^{-1} \sum_{i=1}^n \pi_i^{-1}$.

In Table 4.1, we have the coverage probabilities, the lower and the upper tail error rates, the length averages and length variances of the confidence intervals. The values for large sampling fractions ($N = 150$) are given between brackets. The coverage of the standard approach based on the central limit theorem and linearisation is significantly larger than 95% in all the cases considered. This is due to the fact that the point estimator has a positively skewed sampling distribution. This explains the null upper tail error rate. The linearised variance estimator is also biased. The Woodruff (1952) approach gives confidence intervals which are as good as the empirical likelihood confidence intervals in term of coverage and stability of the confidence intervals, but with lower coverages with small sampling fraction. The rescaled bootstrap confidence intervals may have slightly higher coverage probabilities compared to the other approaches. For small sampling fractions, the performance of the proposed empirical likelihood approach is similar to the rescaled bootstrap. Nevertheless, with large sampling fractions, the rescaled bootstrap confidence intervals may have larger than 95% intended coverage probability, because this approach does not includes finite population correction factors. With large sampling fractions, the direct bootstrap has better coverage because it includes finite population corrections. However, the coverage of the direct bootstrap tends to be smaller than 95%. The direct bootstrap has a low coverage of 89.7% when $\rho(y_i, \hat{y}_i) = 0.80$ and $n = 80$. The proposed approach gives a coverage of 94.2% in this situation. Note that the direct bootstrap has larger variances for the lengths. This means that the direct bootstrap confidence intervals are more volatile than empirical likelihood confidence intervals.

A summary of the results for confidence intervals of quantile $Y_{0.25}$ are presented in the Table 4.2. The outstanding findings are the following: the proposed empirical likelihood, rescaled bootstrap and Woodruff approach are close to the nominal level in case of small sampling fraction except when $n = 80$ and $\rho(y_i, \hat{y}_i) = 0.80$. Empirical likelihood and Woodruff approach confidence intervals tend to perform similarly well and close to the intended nominal coverage probability for large sampling fraction.

A general good performance of the five approaches for the median confidence intervals $Y_{0.50}$ would be expected. However, confidence intervals given in Table 4.3 show only a good behaviour of Woodruff and empirical likelihood approaches: the coverage probabilities are the nearest to the intended nominal level of 95% in large and small sampling fraction cases. The rescaled bootstrap approach is not suitable for large sampling fraction; it leads to confidence interval coverage larger than the intended level, as a consequence of a serious overestimation of the variance.

The Quantile-Quantile plots (a) and (d) in the Figures 4.1, 4.2, 4.3 and 4.4 show that the empirical log-likelihood ratio functions approximate to a chi-square distribution. In despite of the point estimator bias, the coverages are good and have stable performance, except when for large sampling fraction and $n = 80$ $\rho(y_i, \hat{y}_i) = 0.80$ which can be anticipated from Figure 4.4 (d). The Quantile-Quantile plots of the proposed empirical log-likelihood ratio function, (d) in Figures 4.1, 4.2, 4.3 and 4.4, show the best approximations to a chi-square distribution happen for $Y_{0.50}$, the medians. The confidence intervals based on linearisation variance estimator ameliorate its performance. In this case, the coverages are closer to 95% level but still they present an overstated upper tail error.

Table 4.1: Coverages of 95% confidence intervals for the quantile $Y_{0.10}$. The values not in brackets are for the population of size $N = 800$ (small sampling fractions). The values in brackets are for the population of size $N = 150$ (large sampling fractions).

$\rho(y_i, \hat{y}_i)$	n	Approaches	Coverage Prob (%)	Lower Tail Error (%)	Upper Tail Error (%)	Average Length	var Length
0.3	40	Proposed $\mathbf{c}_i = \pi_i$	93.3 (92.8)	4.1 (1.9)	2.6 (5.3)	0.73 (0.67)	0.065 (0.046)
		Direct bootstrap	92.1 (91.7)	4.3 (5.4)	3.6 (2.9)	0.79 (0.72)	0.080 (0.058)
		Woodruff	91.3 (93.5)	4.5 (2.5)	4.2 (4.0)	0.66 (0.66)	0.056 (0.043)
		Rescaled bootstrap	93.9 (95.6)	4.7 (3.0)	1.4 (1.4)	0.77 (0.76)	0.062 (0.045)
		Linearisation	98.9 (99.6)	1.1 (0.4)	0.0 (0.0)	1.17 (1.34)	0.055 (0.049)
0.3	80	Proposed $\mathbf{c}_i = \pi_i$	96.5 (93.6)	0.8 (1.0)	2.7 (5.4)	0.57 (0.43)	0.024 (0.013)
		Direct bootstrap	92.2 (92.3)	4.5 (4.8)	3.3 (2.9)	0.56 (0.44)	0.028 (0.016)
		Woodruff	95.7 (95.1)	1.3 (1.8)	3.0 (3.1)	0.55 (0.44)	0.023 (0.014)
		Rescaled bootstrap	95.4 (98.3)	3.3 (1.3)	1.3 (0.4)	0.57 (0.56)	0.024 (0.018)
		Linearisation	99.4 (99.9)	0.6 (0.1)	0.0 (0.0)	0.86 (0.85)	0.014 (0.007)
0.8	40	Proposed $\mathbf{c}_i = \pi_i$	92.9 (91.8)	4.0 (2.3)	3.1 (5.9)	0.54 (0.37)	0.039 (0.018)
		Direct bootstrap	92.6 (91.3)	4.1 (6.0)	3.3 (2.7)	0.59 (0.39)	0.044 (0.018)
		Woodruff	90.7 (92.7)	4.8 (3.6)	4.5 (3.7)	0.48 (0.36)	0.033 (0.015)
		Rescaled bootstrap	93.9 (94.6)	4.8 (3.8)	1.3 (1.6)	0.57 (0.42)	0.034 (0.020)
		Linearisation	96.5 (99.5)	3.5 (0.5)	0.0 (0.0)	0.73 (0.62)	0.028 (0.012)
0.8	80	Proposed $\mathbf{c}_i = \pi_i$	95.9 (94.2)	1.6 (0.6)	2.5 (5.2)	0.41 (0.23)	0.015 (0.005)
		Direct bootstrap	93.3 (89.7)	3.3 (7.7)	3.4 (2.6)	0.42 (0.23)	0.018 (0.005)
		Woodruff	94.3 (94.3)	2.0 (2.4)	3.7 (3.3)	0.39 (0.22)	0.015 (0.004)
		Rescaled bootstrap	96.7 (98.7)	2.7 (0.8)	0.6 (0.5)	0.42 (0.29)	0.015 (0.005)
		Linearisation	97.8 (99.8)	2.2 (0.2)	0.0 (0.0)	0.55 (0.41)	0.007 (0.002)

Table 4.2: Coverages of 95% confidence intervals for the quantile $Y_{0.25}$. The values not in brackets are for the population of size $N = 800$ (small sampling fractions). The values in brackets are for the population of size $N = 150$ (large sampling fractions).

$\rho(y_i, \hat{y}_i)$	n	Approaches	Coverage Prob (%)	Lower Tail Error (%)	Upper Tail Error (%)	Average Length	var Length
0.30	40	Proposed $\mathbf{c}_i = \pi_i$	94.6 (95.3)	1.7 (1.0)	3.7 (3.7)	1.13 (0.98)	0.115 (0.094)
		Direct bootstrap	92.4 (93.5)	6.0 (6.1)	1.6 (0.4)	1.13 (1.01)	0.137 (0.106)
		Woodruff	94.0 (94.9)	3.2 (2.2)	2.8 (2.9)	1.09 (1.01)	0.115 (0.104)
		Rescaled bootstrap	95.4 (97.4)	2.3 (1.3)	2.3 (1.3)	1.13 (1.12)	0.116 (0.129)
		Linearisation	98.0 (100.0)	1.1 (0.0)	0.9 (0.0)	1.41 (1.70)	0.064 (0.076)
0.30	80	Proposed $\mathbf{c}_i = \pi_i$	95.9 (94.0)	1.6 (1.8)	2.5 (4.2)	0.79 (0.55)	0.038 (0.013)
		Direct bootstrap	92.6 (92.2)	5.6 (5.1)	1.8 (2.7)	0.78 (0.57)	0.038 (0.016)
		Woodruff	94.6 (94.4)	3.1 (3.0)	2.3 (2.6)	0.76 (0.54)	0.037 (0.015)
		Rescaled bootstrap	95.7 (98.8)	2.7 (0.5)	1.6 (0.7)	0.79 (0.74)	0.039 (0.022)
		Linearisation	98.9 (99.9)	0.6 (0.0)	0.5 (0.1)	0.99 (1.00)	0.016 (0.009)
0.80	40	Proposed $\mathbf{c}_i = \pi_i$	95.1 (94.7)	1.7 (1.0)	3.2 (4.3)	0.77 (0.61)	0.039 (0.037)
		Direct bootstrap	91.3 (92.8)	6.1 (5.8)	2.6 (1.4)	0.78 (0.64)	0.048 (0.044)
		Woodruff	94.1 (95.0)	3.5 (2.6)	2.4 (2.4)	0.77 (0.62)	0.041 (0.034)
		Rescaled bootstrap	95.2 (97.3)	2.6 (1.3)	2.2 (1.4)	0.77 (0.70)	0.041 (0.037)
		Linearisation	95.5 (98.1)	1.5 (1.9)	3.0 (0.0)	0.83 (0.78)	0.021 (0.009)
0.80	80	Proposed $\mathbf{c}_i = \pi_i$	96.5 (95.4)	1.5 (0.6)	2.0 (4.0)	0.56 (0.34)	0.013 (0.012)
		Direct bootstrap	93.9 (93.7)	4.5 (6.1)	1.6 (0.2)	0.56 (0.35)	0.017 (0.012)
		Woodruff	96.7 (96.2)	1.9 (1.6)	1.4 (2.2)	0.54 (0.35)	0.013 (0.013)
		Rescaled bootstrap	97.0 (99.8)	1.9 (0.1)	1.1 (0.1)	0.56 (0.49)	0.013 (0.016)
		Linearisation	96.9 (99.5)	0.2 (0.4)	2.9 (0.1)	0.58 (0.47)	0.005 (0.001)

Table 4.3: Coverages of 95% confidence intervals for the quantile $Y_{0.50}$. The values not in brackets are for the population of size $N = 800$ (small sampling fractions). The values in brackets are for the population of size $N = 150$ (large sampling fractions).

$\rho(y_i, \hat{y}_i)$	n	Approaches	Coverage Prob (%)	Lower Tail Error (%)	Upper Tail Error (%)	Average Length	var Length
0.30	40	Proposed $\mathbf{c}_i = \pi_i$	94.1 (94.5)	2.5 (1.4)	3.4 (4.1)	1.82 (2.13)	0.274 (0.499)
		Direct bootstrap	91.3 (92.0)	7.4 (7.3)	1.3 (0.7)	1.84 (2.22)	0.293 (0.544)
		Woodruff	94.6 (95.9)	3.4 (2.5)	2.0 (1.6)	1.87 (2.27)	0.306 (0.556)
		Rescaled bootstrap	94.0 (97.4)	2.7 (1.1)	3.3 (1.5)	1.80 (2.46)	0.274 (0.575)
		Linearisation	91.0 (92.0)	0.5 (1.0)	8.5 (7.0)	1.74 (2.00)	0.159 (0.166)
0.30	80	Proposed $\mathbf{c}_i = \pi_i$	95.6 (95.0)	1.6 (1.6)	2.8 (3.4)	1.31 (1.20)	0.075 (0.056)
		Direct bootstrap	92.1 (93.5)	6.4 (5.5)	1.5 (1.0)	1.30 (1.24)	0.085 (0.062)
		Woodruff	94.7 (95.3)	2.8 (2.9)	2.5 (1.8)	1.28 (1.21)	0.075 (0.065)
		Rescaled bootstrap	95.1 (99.4)	1.7 (0.2)	3.2 (0.4)	1.31 (1.72)	0.075 (0.131)
		Linearisation	90.9 (92.6)	0.7 (0.5)	8.4 (6.9)	1.22 (1.14)	0.034 (0.018)
0.80	40	Proposed $\mathbf{c}_i = \pi_i$	94.6 (94.7)	1.7 (1.2)	3.7 (4.1)	0.96 (0.86)	0.067 (0.059)
		Direct bootstrap	93.2 (95.6)	4.9 (3.2)	1.9 (1.2)	0.96 (0.87)	0.073 (0.064)
		Woodruff	95.2 (96.4)	3.3 (2.5)	1.5 (1.1)	0.98 (0.89)	0.066 (0.060)
		Rescaled bootstrap	94.9 (97.6)	1.7 (0.6)	3.4 (1.8)	0.95 (1.00)	0.066 (0.065)
		Linearisation	94.2 (94.4)	2.0 (1.3)	3.8 (4.3)	0.94 (0.87)	0.037 (0.018)
0.80	80	Proposed $\mathbf{c}_i = \pi_i$	96.0 (95.2)	1.7 (1.1)	2.3 (3.7)	0.65 (0.45)	0.020 (0.010)
		Direct bootstrap	93.2 (97.3)	4.2 (2.3)	2.6 (0.4)	0.64 (0.44)	0.021 (0.010)
		Woodruff	95.0 (95.8)	3.6 (2.2)	1.4 (2.0)	0.63 (0.45)	0.018 (0.011)
		Rescaled bootstrap	95.9 (99.9)	1.8 (0.0)	2.3 (0.1)	0.65 (0.68)	0.020 (0.018)
		Linearisation	95.8 (96.7)	1.6 (0.1)	2.6 (3.2)	0.64 (0.50)	0.008 (0.002)

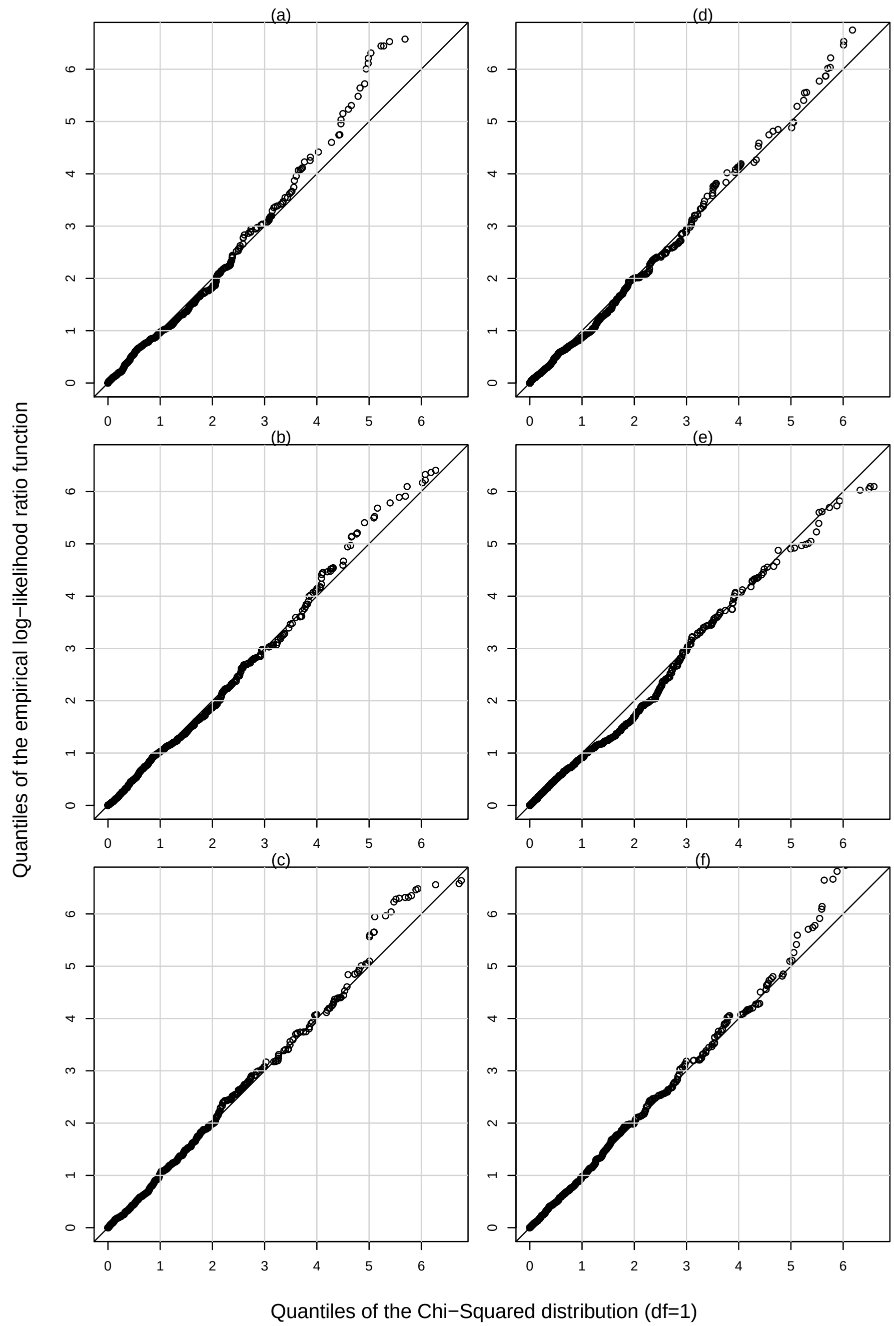

Figure 4.1: Quantile-Quantile plot of the observed distribution of the proposed empirical log-likelihood ratio function $\widehat{r}(\theta_0)$ when $\boldsymbol{c}_i = \pi_i,\ N = 800,\ n = 40$. The parameters of interest θ_0 are the quantiles (a)$Y_{0.10}$, (b)$Y_{0.25}$, (c)$Y_{0.50}$ with $\rho\left(y_i, \hat{y}_i\right) = 0.30$; (d)$Y_{0.10}$, (e)$Y_{0.25}$, (f)$Y_{0.50}$ with $\rho\left(y_i, \hat{y}_i\right) = 0.80$. We considered that we have a negligible sampling fraction. The approach of Section 4.2 is used. The data are generated with the model (3.9.1) with $\beta = 0$.

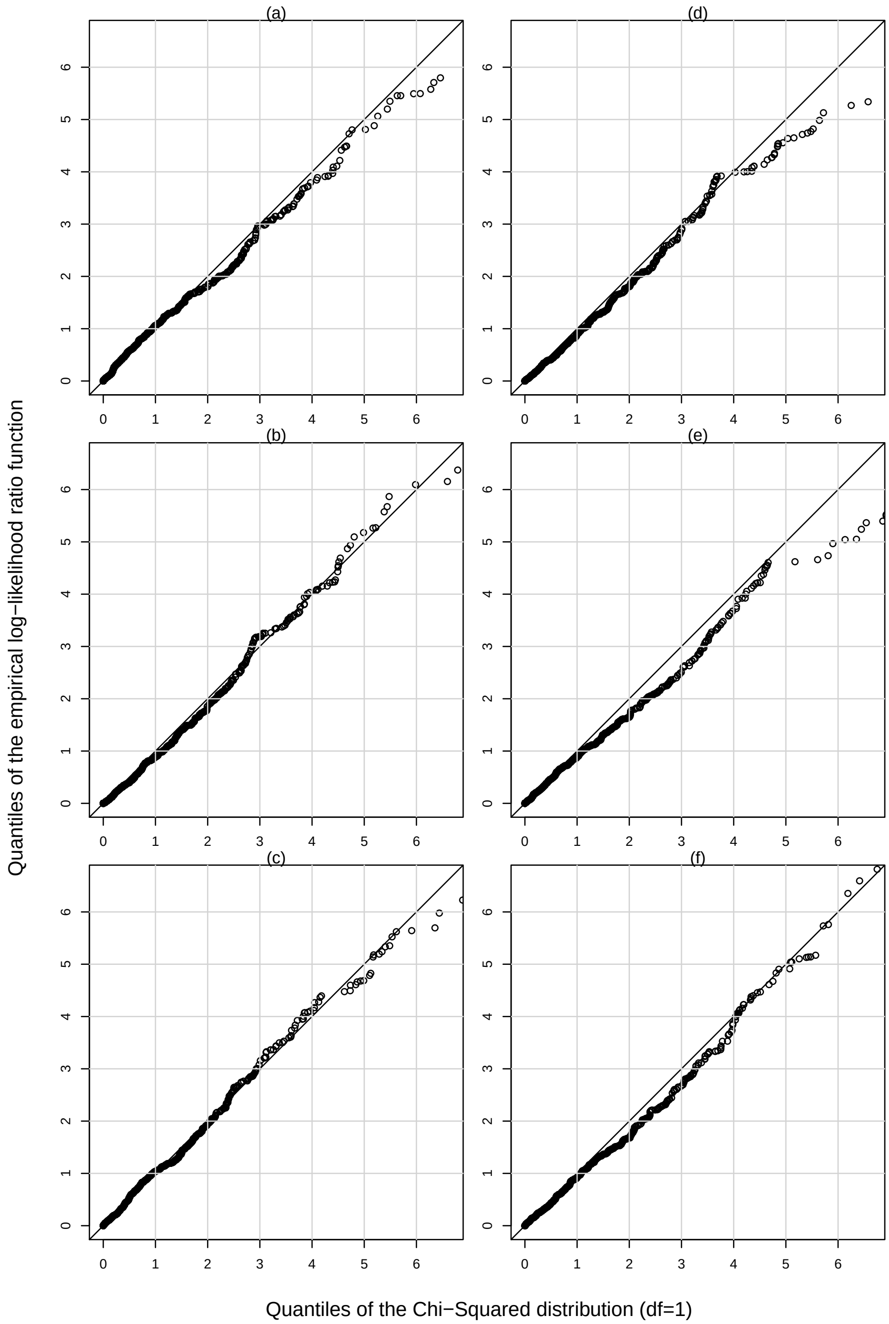

Figure 4.2: Quantile-Quantile plot of the observed distribution of the proposed empirical log-likelihood ratio function $\widehat{r}(\theta_0)$ when $\boldsymbol{c}_i = \pi_i,\ N = 800,\ n = 80$. The parameters of interest θ_0 are the quantiles (a)$Y_{0.10}$, (b)$Y_{0.25}$, (c)$Y_{0.50}$ with $\rho\left(y_i, \hat{y}_i\right) = 0.30$; (d)$Y_{0.10}$, (e)$Y_{0.25}$, (f)$Y_{0.50}$ with $\rho\left(y_i, \hat{y}_i\right) = 0.80$. We considered that we have a negligible sampling fraction. The approach of Section 4.2 is used. The data are generated with the model (3.9.1) with $\beta = 0$.

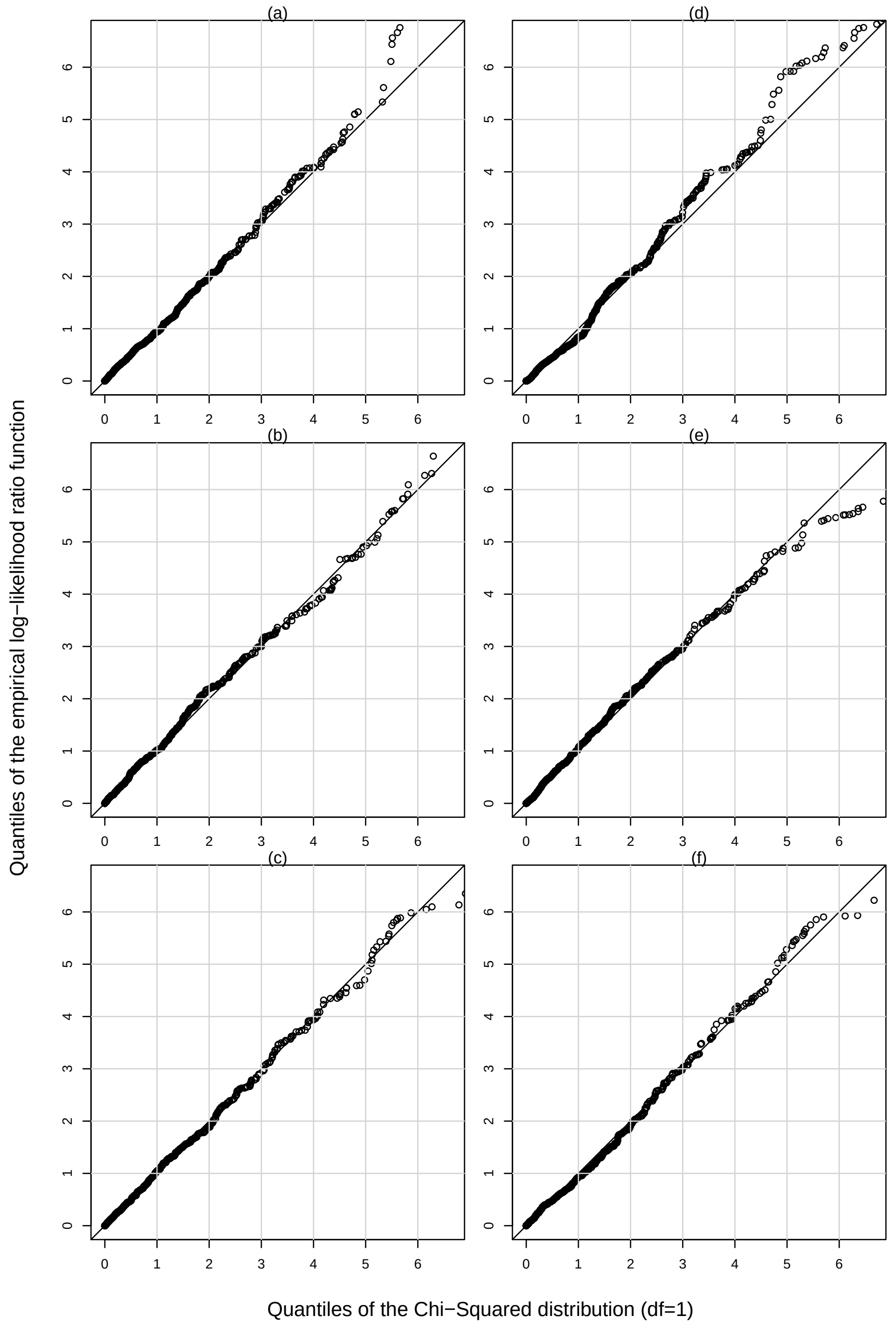

Figure 4.3: Quantile-Quantile plot of the observed distribution of the proposed empirical log-likelihood ratio function $\widehat{r}(\theta_0)$ when $\boldsymbol{c}_i = \pi_i, \ N = 150, \ n = 40$. The parameters of interest θ_0 are the quantiles (a)$Y_{0.10}$, (b)$Y_{0.25}$, (c)$Y_{0.50}$ with $\rho\left(y_i, \hat{y}_i\right) = 0.30$; (d)$Y_{0.10}$, (e)$Y_{0.25}$, (f)$Y_{0.50}$ with $\rho\left(y_i, \hat{y}_i\right) = 0.80$. We considered that we have a non-negligible sampling fraction. The approach of Section 4.2 is used. The data are generated with the model (3.9.1) with $\beta = 0$.

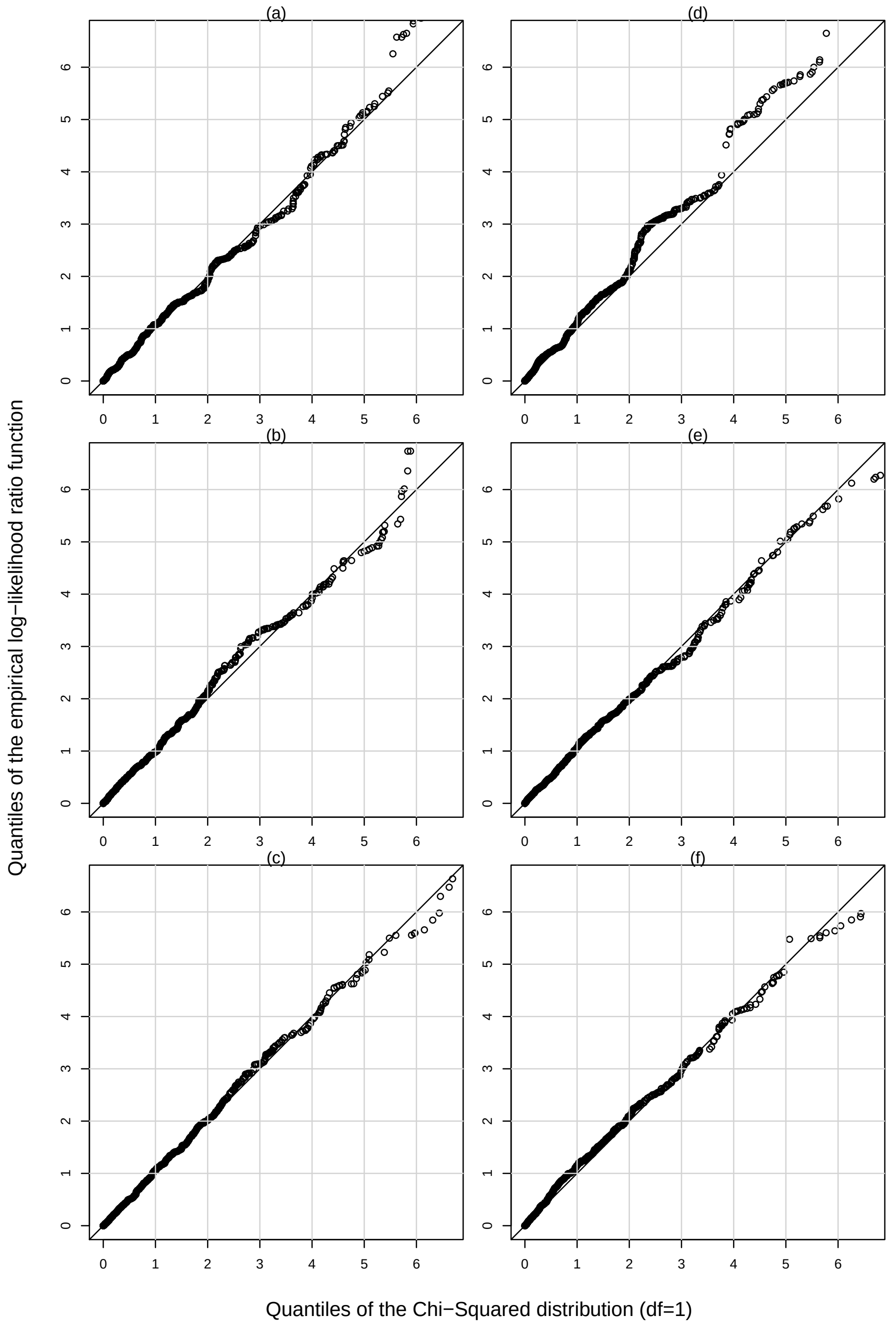

Figure 4.4: Quantile-Quantile plot of the observed distribution of the proposed empirical log-likelihood ratio function $\widehat{r}(\theta_0)$ when $\boldsymbol{c}_i = \pi_i,\ N = 150,\ n = 80$. The parameters of interest θ_0 are the quantiles (a)$Y_{0.10}$, (b)$Y_{0.25}$, (c)$Y_{0.50}$ with $\rho\left(y_i, \hat{y}_i\right) = 0.30$; (d)$Y_{0.10}$, (e)$Y_{0.25}$, (f)$Y_{0.50}$ with $\rho\left(y_i, \hat{y}_i\right) = 0.80$. We considered that we have a non-negligible sampling fraction. The approach of Section 4.2 is used. The data are generated with the model (3.9.1) with $\beta = 0$.

Two additional simulation studies were run with the sugar cane farm data (Chambers and Dunstan, 1986) and the data of Swedish municipalities MU284 (Särndal et al., 1992) for estimation of confidence intervals of quantiles Y_q when $q = 0.10$, 0.25 and 0.50. We keep the same simulation settings defined at the beginning of this section. Inclusion probabilities are proportional to the auxiliary variable x_i. In both studies considered sampling fraction were greater than 10%. We considered a strong correlation value $\rho\left(y_i, x_i\right) = 0.90$ where the variable of interest y_i is the harvest sugarcane and the auxiliary variable x_i is area assigned for cane planting in the case of sugar cane farm data, $N = 338$. A moderate correlation, $\rho\left(y_i, x_i\right) = 0.47$, where y_i is the REV84 real estate values according to 1984 assessment (in millions of kronor) and the auxiliary variable x_i is SS82 the number of Social-Democratic seats in municipal council $(N = 284)$, is considered for the Swedish municipalities. Ties in both data sets were treated using the procedure suggested by Owen (2001).

A summary of the results for confidence intervals for the quantiles Y_q of sugar cane farm data are given in Table 4.4. This is an example of strong correlation between y_i and x_i, where the five approaches perform very poorly when $n = 40$ and $Y_{0.10}$. The inflated length of the proposed empirical likelihood approach confidence intervals tend to overstate the coverage probability. In this case none of the approaches shows an acceptable general performance. Woodruff intervals have the confidence intervals with shortest length, due to a strong correlation between the inclusion probabilities and the auxiliary variable when produces a small variance of the Horvitz-Thompson estimator. Although its coverage probabilities are not entirely satisfactory.

In Table 4.5 are summarised the result for the confidence interval of the quantiles Y_q for the Swedish municipalities MU284 data. The proposed empirical likelihood intervals produce somewhat stable result when $n = 40$. Actual coverage probabilities are close to the nominal 95% level. On the other hand, when $n = 80$, the total error rates are overstated higher than 95%. The Woodruff method-based confidence intervals performed satisfactory well in most cases except when $n = 40$ and $Y_{0.10}$ in terms of the actual coverage probability and confidence intervals length. The method based on the linearisation variance estimator produces the longest interval and therefore they performed very poorly with substantially overstated actual coverage probabilities. Resampling methods tend to generate results that to some extent are unstable. However, the lengths of their confidence intervals are in between the length of Woodruff approach and the proposed approach.

Table 4.4: Coverages of 95% confidence intervals for the quantiles $Y_{0.10}, Y_{0.25}, Y_{0.50}$ of y_i = harvest sugarcane, auxiliary variable x_i = area assigned for cane planting (Chambers and Dunstan, 1986), $N = 338, \rho(y_i, x_i) = 0.90$.

n	Approaches	quantile	Coverage Prob (%)	Lower Tail Error (%)	Upper Tail Error (%)	Average Length	var Lenght
40	Proposed $\mathbf{c}_i = \pi_i$	0.10	86.8	9.1	4.1	1134.1	231110.4
	Direct bootstrap		87.4	5.6	7.0	1218.1	227298.1
	Woodruff		81.6	13.9	4.5	921.0	133454.1
	Rescaled bootstrap		85.4	13.7	0.9	1102.0	139114.1
	Linearisation		92.7	4.2	3.1	1410.1	172045.5
	Proposed $\mathbf{c}_i = \pi_i$	0.25	95.3	1.9	2.8	1427.8	206447.8
	Direct bootstrap		93.0	3.1	3.9	1335.8	184734.0
	Woodruff		93.0	5.7	1.3	1324.4	184407.9
	Rescaled bootstrap		95.1	3.6	1.3	1367.9	171805.3
	Linearisation		96.2	2.4	1.4	1382.2	62208.7
	Proposed $\mathbf{c}_i = \pi_i$	0.50	96.1	1.4	2.5	1530.7	159041.5
	Direct bootstrap		94.6	3.0	2.4	1510.3	171618.5
	Woodruff		94.6	4.9	0.5	1514.3	153953.7
	Rescaled bootstrap		96.2	1.9	1.9	1530.5	162062.3
	Linearisation		97.3	1.3	1.4	1537.1	53809.2
80	Proposed $\mathbf{c}_i = \pi_i$	0.10	94.9	1.7	3.4	1081.0	139455.5
	Direct bootstrap		93.8	2.1	4.1	997.8	131267.4
	Woodruff		92.8	4.0	3.2	982.3	118787.5
	Rescaled bootstrap		95.6	3.3	1.1	1043.6	117513.8
	Linearisation		94.4	3.1	2.5	981.1	33255.2
	Proposed $\mathbf{c}_i = \pi_i$	0.25	96.4	1.0	2.6	982.2	58974.6
	Direct bootstrap		93.2	2.1	4.7	912.2	61816.1
	Woodruff		95.2	3.5	1.3	877.8	52122.9
	Rescaled bootstrap		96.9	1.6	1.5	959.7	54989.6
	Linearisation		95.6	1.1	3.3	913.0	10658.7
	Proposed $\mathbf{c}_i = \pi_i$	0.50	97.8	0.6	1.6	1057.5	44430.9
	Direct bootstrap		93.6	3.3	3.1	980.0	48963.8
	Woodruff		95.3	4.0	0.7	957.4	39762.5
	Rescaled bootstrap		97.5	0.8	1.7	1055.0	46634.6
	Linearisation		96.7	1.4	1.9	1031.3	8245.5

Table 4.5: Coverages of 95% confidence intervals for the quantiles $Y_{0.10}, Y_{0.25}, Y_{0.50}$ of $y_i = \text{REV}84$ real estate values according to 1984 assessment (in millions of kronor), auxiliary variable $x_i = \text{SS}82$ number of Social-Democratic seats in municipal council (Särndal et al., 1992), $N = 284$, $\rho(y_i, x_i) = 0.47$.

n	Approaches	quantile	Coverage Prob (%)	Lower Tail Error (%)	Upper Tail Error (%)	Average Length	var Lenght
40	Proposed $\mathbf{c}_i = \pi_i$	0.10	95.5	1.7	2.8	538.3	33895.3
	Direct bootstrap		95.4	2.6	2.0	513.3	34093.3
	Woodruff		93.0	2.6	4.4	484.5	31193.2
	Rescaled bootstrap		96.0	3.1	0.9	552.5	28570.5
	Linearisation		99.2	0.7	0.1	1059.2	214416.1
	Proposed $\mathbf{c}_i = \pi_i$	0.25	96.0	1.6	2.4	727.1	46572.2
	Direct bootstrap		91.1	1.1	7.8	703.3	46997.9
	Woodruff		94.8	3.1	2.1	691.6	44991.1
	Rescaled bootstrap		95.9	2.4	1.7	733.2	50443.7
	Linearisation		99.7	0.2	0.1	1490.6	438134.7
	Proposed $\mathbf{c}_i = \pi_i$	0.50	95.3	1.4	3.3	1286.4	152969.4
	Direct bootstrap		91.4	1.4	7.2	1258.7	155414.7
	Woodruff		94.7	3.4	1.9	1281.9	153775.9
	Rescaled bootstrap		94.9	1.6	3.5	1278.5	149108.2
	Linearisation		97.6	0.4	2.0	1860.0	666237.6
80	Proposed $\mathbf{c}_i = \pi_i$	0.10	96.5	1.8	1.7	380.8	15503.5
	Direct bootstrap		95.4	2.7	1.9	317.6	11932.7
	Woodruff		94.7	2.9	2.4	318.0	13559.6
	Rescaled bootstrap		97.1	2.5	0.4	380.5	14099.7
	Linearisation		99.5	0.5	0.0	791.3	55630.3
	Proposed $\mathbf{c}_i = \pi_i$	0.25	96.9	1.4	1.7	521.2	11796.6
	Direct bootstrap		91.8	2.4	5.8	473.5	12831.6
	Woodruff		95.0	3.1	1.9	461.8	11217.7
	Rescaled bootstrap		96.8	2.0	1.2	519.2	12343.3
	Linearisation		100.0	0.0	0.0	1059.9	92573.6
	Proposed $\mathbf{c}_i = \pi_i$	0.50	97.3	1.1	1.6	924.1	32833.3
	Direct bootstrap		93.0	2.3	4.7	844.5	38323.1
	Woodruff		94.3	4.0	1.7	834.1	28115.0
	Rescaled bootstrap		97.4	1.0	1.6	924.4	33179.7
	Linearisation		98.5	0.7	0.8	1284.5	127123.7

4.4 Concluding remarks

Standard confidence intervals based upon the central limit theorem can perform poorly when the sampling distribution is not normal. The coverage and the tail errors can be also different from their intended levels. On the other hand, empirical likelihood confidence intervals may be better in this situation, as empirical likelihood confidence intervals are determined by the distribution of the data and the range of the parameter space is preserved. Note that the distribution of a point estimator of is not necessarily normal, and the proposed empirical likelihood approach does not rely on the normality of the point estimator. Standard confidence intervals based on the central limit theorem require normality and variance estimates which often involve linearisation or resampling. The proposed method does not rely on normality, variance estimates, linearisation or resampling, even if the parameter of interest is not linear. Empirical likelihood confidence intervals can be easier to compute than standard confidence intervals based on variance estimates. It provides an alternative to the bootstrap, when linearisation cannot be used. The proposed approach has some advantages over the bootstrap approach. It is less computationally intensive than the bootstrap. Our simulations study also show that bootstrap confidence intervals may not have the right coverage and may be more unstable. When the sample size is large, the Woodruff approach gives confidence intervals which are as good as the empirical likelihood confidence intervals in term of coverage and stability of the confidence intervals. However, the Woodruff approach relies on variance estimates and joint inclusion probabilities. Furthermore, this approach is only designed for quantiles. The empirical likelihood approach can be used for a wider class of point estimators. The inadequacy of the variance estimation based on linerisation for generating confidence intervals for population quantiles (Kovar et al., 1988) was confirmed. The results obtained for the proposed empirical likelihood approach in the case of the sugar cane farms and Swedish municipalities MU284 data imply that some of regularity conditions (see Section 3.3) do not hold. The procedure to break up ties in discrete data is an additional source of randomness that may been affecting the estimation of confidence intervals.

5

Empirical Likelihood confidence intervals for the persistent-risk-of-poverty rate

5.1 Introduction

This chapter shows how the proposed empirical likelihood approach can be used to compute confidence intervals for *persistent-risk-of-poverty rate* indicator, one of the set of *Laeken* indicators; this set of 18 complex indicators of poverty and inequality was conceived in the framework of the *European Statistics on Income and Living Conditions* (EU-SILC) project which is the main instrument for the compilation of comparable indicators of social cohesion in the European Union (Osier, 2009). Every year from 2003, information on income, poverty and social inclusion are collected from approximately 300,000 households across Europe. In order to monitor poverty across the European Union, several poverty indicator are estimated. The *persistent-risk-of-poverty* rate indicator is the core EU-SILC longitudinal indicator. For a four-year panel, it is defined as the share of persons who are at-risk-of-poverty at the fourth wave of the panel and at two of the three preceding waves. The *at-the-risk-of-poverty* threshold is set at 60% of the national median equivalised disposable income after social transfers (Eurostat, 2012). The *persistent-risk-of-poverty* rate can be decomposed according to household or personal characteristics (e.g. age group, gender, geographical region or most activity status). However, the *at-the-risk-of-poverty* threshold was calculated over the population total and not over the population breakdown or partition which is considered (Osier, 2009). We use EU-SILC survey user database from 2009.

5.2 Empirical confidence intervals for persistent-risk-of-poverty rate

We adopted the approach described in Section 3.4 to construct the 95% empirical likelihood confidence intervals. The estimating function $g_i(\theta)$ of equation (3.1.1) for the persistent-risk-of-poverty rate is defined as $g_i(\theta) = y_i - \theta_0$ where $y_i = 1$ if the individual i has a persistent-risk-of-poverty and $y_i = 0$ otherwise.

Often the estimates in a survey are also intended for subpopulations or domains. The simplest way to define domain is as a subpopulation for which separate point estimates and confidence intervals are required before or after the survey planning stage. The domain size is usually random. The domains form a disjoint partition of a population into subsets. The introduction of the domain indicator membership function δ_{di} serves to divide the population U into D domains $U_1, \ldots, U_d, \ldots, U_D$, where $\delta_{di} = 1$ if the i unit belongs to U_d and $\delta_{di} = 0$ otherwise (Särndal et al., 1992). The domain size is $\sum_{i \in U} \delta_{di} = N_d$.

The *maximum empirical likelihood estimate* $\widehat{\theta}$ of θ_0 is defined by the value of θ which minimises the function $\widehat{r}(\theta)$, defined by (3.3.1). As $\widehat{r}(\theta)$ is a positive function with a minimum value of zero, $\widehat{\theta}$ is the solution of $\widehat{r}(\theta) = 0$.

If the parameter of interest is the persistent-risk-of-poverty for the domain d, we use

$$g_i(\theta) = (y_i - \theta_0)\delta_{di}. \qquad (5.2.1)$$

When θ_0 is the persistent-risk-of-poverty for a domain d, the $g_i(\theta)$ are given by (5.2.1) and

$$\widehat{\theta} = \frac{\sum_{i=1}^{n} y\delta_{di}\pi_i^{-1}}{\sum_{i=1}^{n} \delta_{di}\pi_i^{-1}} \tag{5.2.2}$$

which is the standard Hájek (1971) estimator of the persistent-risk-of-poverty.

When $g_i(\theta)$ is given by (5.2.1), the interval (3.5.3) gives a confidence interval for the persistent-risk-of-poverty, if we have a single stage stratified sampling design. Most of the EU-SILC surveys are multi-stage cluster surveys. Therefore, the approach described in Section 3.4 cannot be directly implemented. We proposed to use an ultimate cluster approach which consists in treating the primary sampling unit (PSU) totals estimates as a response variable and the PSU as sampling units. The inclusion probabilities are the first-order inclusion of the PSUs and the stratification used is at PSU level. Assuming that the sub-sample is a probability sample selected in the PSU independently of the selection of a sub-sample in another PSU, the variance estimation is consistent if there are at least two sampled elements at each stage, or substratum if a stratified design were applied in the first-stage units (Hansen et al., 1953). Within each PSU, sub-sampling of any complexity may be involved. It is also necessary to have small sampling fractions. These conditions are usually met with social surveys. However, there are exceptions. For example, the EU-SILC survey for Belgium uses large sampling fractions, as the number of clusters in the population is not large. If these conditions hold, the variance estimator (3.6.2) is approximately unbiased and $\widehat{r}(\theta_0)$ asymptotically a chi-squared distribution with one degree of freedom. The persistent-risk-of-poverty rate estimation contains sampling errors due to the estimation of the poverty thresholds. For simplicity, we assume that the poverty threshold is fixed which ensures conservative confidence intervals (Berger and Skinner, 2003).

Calibration and stratification variables were not available. We used the Nomenclature of Territorial Units for Statistics (NUTS2) geographical region as a proxy for stratification and the effect of calibration was ignored. The confidence intervals for the persistent-risk-of-poverty rate are estimated for the overall population and domains defined by gender and age groups. We adopted the approach described in Section 3.4 to construct confidence intervals, treating the sample of PSUs as a single stage sample. Standard confidence intervals and rescaled bootstrap confidence intervals (Rao and Wu, 1988; Rao et al., 1992) were compute in order to allow the comparison with other available methods. The former assumes normal distributed data for its variance estimation; the latter is based on resampling techniques.

The results are given in Table 5.1. The empirical likelihood confidence intervals are consistent with the standard confidence intervals. The sampling distribution must be approximately normal, as we do not observe large differences between these intervals. Nevertheless, the bounds of the empirical likelihood intervals are always larger than the bounds of the standard intervals and the bootstrap intervals tend to be located between empirical likelihood intervals and standard interval. This difference is particularly pronounced for Latvia. This is probably due to a slight skewness of the sampling distribution and based on the three approaches evaluated here there is no evidence of outstanding differences. The distribution of the persistent-risk-of-poverty rate, therefore, should be approximately normal.

Table 5.1: Estimates of persistent-risk-of-poverty rate for 2009. 95% empirical likelihood confidence intervals. The standard confidence intervals are given between brackets. Rescaled bootstrap confidence intervals are given between squared brackets.

Country	Rate %	Lower %	Upper %
Austria	5.9	4.6 (4.4) [4.5]	7.7 (7.5) [7.6]
Belgium	8.9	7.1 (6.9) [7.0]	11.3(11.0)[11.1]
Bulgaria	10.7	8.2 (7.9) [8.0]	13.9(13.4)[13.6]
Cyprus	11.0	9.1 (8.9) [9.1]	13.1(13.0)[12.9]
Czech Republic	3.6	2.5 (2.3) [2.4]	5.1 (4.8) [5.0]
Denmark	6.3	4.6 (4.4) [4.5]	8.4 (8.2) [8.2]
Estonia	13.0	11.1(10.9)[11.1]	15.1(15.0)[15.1]
Spain	11.1	9.5 (9.4) [9.4]	12.9(12.8)[12.8]
Finland	6.9	5.5 (5.4) [5.5]	8.6 (8.4) [8.3]
France	6.9	6.0 (6.0) [5.9]	7.9 (7.8) [8.0]
Greece	14.5	11.8(11.8)[11.5]	17.8(17.2)[17.6]
Hungary	8.3	6.6 (6.5) [6.5]	10.4(10.1)[10.3]
Ireland	6.3	4.3 (3.7) [4.1]	9.5 (8.9) [8.9]
Iceland	4.2	2.4 (2.0) [2.0]	6.9 (6.4) [6.7]
Italy	13.4	11.2(11.3)[10.9]	16.0(15.5)[16.1]
Lithuania	11.7	9.2 (8.9) [9.1]	15.2(14.5)[15.0]
Luxembourg	8.8	6.9 (6.7) [6.9]	11.5(11.0)[11.2]
Latvia	17.7	14.0(13.5)[13.6]	23.9(22.0)[23.1]
Malta	6.2	4.6 (4.4) [4.5]	8.2 (8.0) [8.1]
Netherlands	6.4	4.1 (3.7) [3.7]	9.8 (9.0) [9.4]
Norway	5.4	4.2 (4.0) [4.1]	6.9 (6.7) [6.8]
Poland	10.2	7.5 (8.6) [7.4]	13.4(11.7)[13.4]
Portugal	10.0	7.8 (7.6) [7.7]	12.7(12.4)[12.5]
Sweden	5.7	4.4 (4.2) [4.2]	7.2 (7.1) [7.0]
Slovakia	5.0	3.5 (3.3) [3.4]	7.1 (6.7) [6.9]
United Kingdom	8.4	6.7 (6.5) [6.5]	10.5(10.2)[10.4]

5.3 Estimation of confidence intervals for persistent-risk-of-poverty rate for domains

We consider several domains of interest by cross-classifying gender (male and female) and three age groups: (i) 16 year-old to 25 year-old; (ii) 25 year-old to 44 year-old and (iii) older than 44 year-old. This produces six domains of study.

Table 5.2 gives the results for males and females. These results show the same pattern as confidence intervals for the overall persistent-risk-of-poverty rate, under the empirical likelihood, standard and rescaled bootstrap approaches. Ireland, Lithuania and Latvia showed the most outstanding differences, particularly for the females domains, and Netherlands and Poland for males domains, in the estimation of the lower and upper bounds.

Consider the following two domains: males and females younger than 25 years old (see Table 5.3). The pattern discovered in the previous section remained here. Latvia exhibited the highest differences for the estimation of upper bound between the three approaches. The males domain in Iceland is an example of standard confidence interval drawbacks: the lower bound is negative. Poland's rate estimation lower bound under the standard approach was less than the confidence interval lower bound of the empirical likelihood and rescaled bootstrap approaches.

The results for the age group from 25 year-old to 44 year-old are given in Table 5.4. Note that the bounds of the standard intervals for males are negative for Ireland, Austria, Malta and Denmark. The bootstrap bounds and the empirical likelihood bounds are larger than the bounds of the standard intervals. These differences are more pronounced for Austria, Malta, Denmark, the Netherlands, Estonia, Latvia and Greece. This is due to the skewness of the sampling distribution. For Poland, the standard confidence interval is smaller than the other confidence intervals. The lower bound of standard confidence interval for female's rate in Netherlands is negative; Ireland and Lithuania presented the largest differences between the three confidence intervals. The bootstrap bounds tends to be smaller than the bounds of the empirical likelihood intervals. This is probably due to the fact that with the bootstrap the lower tail coverage tends to be too low and the upper tail coverage tends to be too large.

In Table 5.5, we consider the age group above 44 year-old. We have that Iceland has a negative lower bound of the standard confidence interval. Finland, Denmark, Greece and Latvia have the highest difference, particularly in the upper bound for males. An atypical case of the standard confidence lower bound for females is verified in Poland. This interval is shorter than the empirical likelihood and bootstrap confidence intervals.

These confidence intervals are for illustrative purpose only, and are not part of any results officially released by Eurostat. The quality of these confidence intervals relies on the availability and quality of the design variables. These confidence intervals are likely to be conservative because the effect of calibration adjustment was not taken into account. This effect may be more pronounced for Scandinavian countries which rely on heavily on calibration adjustment. The confidence intervals for Belgium might be too short because the sampling fraction is not negligible for Belgium.

Table 5.2: Estimates of at-persistent-risk-of-poverty rate for 2009 by gender. 95% empirical likelihood confidence intervals. The standard confidence intervals are given between brackets. Rescaled bootstrap confidence intervals are given between squared brackets.

Country	Rate %	Lower %	Upper %	Country	Rate %	Lower %	Upper %
Austria				Iceland			
Males	4.6	3.1 (2.9) [3.0]	6.9 (6.4) [6.6]	Males	3.3	1.6 (1.2) [1.2]	6.1 (5.4) [5.5]
Females	7.1	5.5 (5.2) [5.3]	9.3 (9.0) [8.9]	Females	5.1	2.7 (2.2) [2.4]	8.6 (7.9) [8.1]
Belgium				Italy			
Males	7.8	5.9 (5.7) [5.8]	10.1 (9.8) [9.8]	Males	12.3	10(10.0) [9.9]	15.1(14.5)[14.9]
Females	10	7.8 (7.5) [7.5]	13.1(12.5)[12.9]	Females	14.5	12.1(12.3)[11.9]	17.1(16.6)[17.2]
Bulgaria				Lithuania			
Males	9.7	7.1 (6.8) [6.9]	13.0(12.6)[12.5]	Males	9.4	6.8 (6.5) [6.7]	12.9(12.3)[12.6]
Females	11.6	8.9 (8.6) [8.8]	15.1(14.5)[14.9]	Females	13.7	10.8(10.5)[10.6]	17.9(16.9)[17.6]
Cyprus				Luxembourg			
Males	8.3	6.6 (6.4) [6.6]	10.3(10.2)[10.2]	Males	7.8	5.8 (5.5) [5.6]	10.6(10.0)[10.0]
Females	13.5	11.2(11.0)[11.2]	16.1(15.9)[15.8]	Females	9.9	7.6 (7.3) [7.4]	12.9(12.3)[12.6]
Czech Republic				Latvia			
Males	3.1	2.0 (1.6) [1.8]	5.3 (4.6) [4.9]	Males	15.1	11(10.6)[10.5]	21.3(19.6)[20.4]
Females	3.9	2.9 (2.8) [2.8]	5.5 (5.1) [5.3]	Females	20.0	15.9(15.3)[15.6]	26.9(24.6)[25.8]
Denmark				Malta			
Males	5.9	3.9 (3.6) [3.5]	8.6 (8.1) [8.3]	Males	5.5	3.8 (3.5) [3.6]	8.1 (7.6) [7.7]
Females	6.7	4.6 (4.3) [4.3]	9.5 (9.0) [9.0]	Females	6.9	5.1 (4.9) [5.0]	9.2 (8.9) [9.0]
Estonia				Netherlands			
Males	11.6	9.3 (9.0) [9.3]	14.5(14.1)[13.9]	Males	7.4	4.4 (3.8) [4.0]	12.4(11.0)[11.7]
Females	14.1	11.9(11.8)[11.8]	16.6(16.5)[16.6]	Females	5.4	3.2 (2.8) [3.0]	8.7 (8.0) [8.4]
Spain				Norway			
Males	10.5	8.8 (8.7) [8.6]	12.5(12.3)[12.3]	Males	3.8	2.7 (2.5) [2.7]	5.3 (5.0) [5.2]
Females	12.6	10.9(10.8)[10.7]	14.6(14.5)[14.5]	Females	6.9	4.9 (4.7) [4.8]	9.6 (9.0) [9.3]
Finland				Poland			
Males	5.5	4.1 (3.9) [4.1]	7.5 (7.2) [7.0]	Males	10.3	7.5 (8.6) [7.5]	13.7(11.9)[13.4]
Females	8.2	6.4 (6.2) [6.1]	10.5(10.3)[10.3]	Females	10.1	7.5 (8.5) [7.3]	13.3(11.6)[13.1]
France				Portugal			
Males	6.2	5.3 (5.2) [5.1]	7.3 (7.2) [7.3]	Males	9.4	7.1 (6.8) [7.0]	12.4(12.0)[12.3]
Females	7.6	6.6 (6.6) [6.5]	8.7 (8.6) [8.7]	Females	10.5	8.2 (7.9) [8.0]	13.6(13.1)[13.2]
Greece				Sweden			
Males	13.9	10.9(10.8)[10.7]	18.1(17.1)[17.5]	Males	4.5	3.3 (3.1) [3.2]	6.1 (5.9) [5.9]
Females	15.1	12.2(12.2)[12.1]	18.4(17.9)[18.1]	Females	6.8	5.1 (4.9) [5.0]	8.8 (8.6) [8.7]
Hungary				Slovakia			
Males	8.8	7.0 (6.8) [6.7]	11.2(10.9)[10.8]	Males	4.8	3.0 (2.6) [2.7]	7.5 (6.9) [6.9]
Females	7.8	6.2 (6.0) [6.0]	9.9 (9.6) [9.7]	Females	5.3	3.8 (3.6) [3.7]	7.1 (6.9) [6.9]
Ireland				United Kingdom			
Males	5.9	3.8 (3.3) [3.6]	8.8 (8.4) [8.4]	Males	8.0	6.2 (6.0) [6.1]	10.2(10.0) [9.9]
Females	6.8	4.5 (3.7) [4.2]	10.7 (9.8) [9.8]	Females	8.7	6.8 (6.6) [6.7]	11.2(10.8)[11.1]

Table 5.3: Estimates of at-persistent-risk-of-poverty rate for 2009 by gender and age group $<$ 25 years old. 95% empirical likelihood confidence intervals. The standard confidence intervals are given between brackets. Rescaled bootstrap confidence intervals are given between squared brackets.

Country	Rate %	Lower %	Upper %	Country	Rate %	Lower %	Upper %
Austria				Iceland			
Males	2.1	0.7 (0.2) [0.3]	5.0 (4.2) [4.4]	Males	3.3	0.9 (-0.2) [0.4]	8.4 (6.8) [7.2]
Females	3.9	1.4 (0.5) [0.8]	8.7 (7.3) [7.9]	Females	6.7	3.0 (2.1) [2.3]	12.7(11.3)[12.0]
Belgium				Italy			
Males	8.9	5.6 (4.9) [5.2]	13.6(12.8)[12.8]	Males	16.5	12.3(12.1)[12.0]	22.1(21.0)[22.1]
Females	9.2	4.7 (3.5) [3.9]	17.6(15.0)[15.7]	Females	17.9	13.4(13.9)[13.2]	23.2(21.8)[23.2]
Bulgaria				Lithuania			
Males	15.6	10.1 (9.8) [9.3]	23.1(21.5)[22.5]	Males	11.1	6.0 (5.1) [5.3]	19.4(17.1)[18.6]
Females	11.2	6.5 (5.7) [5.9]	18.2(16.7)[17.6]	Females	15.4	9.2 (8.5) [8.2]	25.2(22.4)[24.1]
Cyprus				Luxembourg			
Males	4.5	2.4 (1.8) [2.1]	7.8 (7.2) [7.3]	Males	13.2	9.1 (8.4) [8.8]	19.6(18.0)[19.0]
Females	7.3	4.2 (3.6) [3.8]	11.9(11.0)[11.5]	Females	15.3	10.5 (9.9)[10.0]	22.5(20.6)[21.4]
Czech Republic				Latvia			
Males	6.2	3.3 (2.4) [2.9]	12.3(10.1)[10.9]	Males	17.6	9.8 (8.7) [7.9]	33.4(26.4)[31.1]
Females	4.8	2.6 (1.9) [2.1]	9.1 (7.8) [8.2]	Females	21.2	11.9 (9.8)[10.3]	42.5(32.6)[36.6]
Denmark				Malta			
Males	3.0	1.1 (0.4) [0.6]	6.7 (5.6) [6.0]	Males	6.0	3.3 (2.6) [3.0]	10.3 (9.4) [9.6]
Females	3.4	1.1 (0.0) [0.3]	8.4 (6.8) [7.0]	Females	7.4	4.3 (3.7) [3.9]	12.0(11.2)[11.5]
Estonia				Netherlands			
Males	13.9	9.6 (9.0) [9.0]	20.4(18.8)[19.6]	Males	12.5	5.4 (3.9) [4.3]	25.5(21.0)[23.8]
Females	10.4	7.2 (6.7) [6.8]	15.0(14.2)[14.5]	Females	5.4	1.8 (0.5) [1.1]	12.6(10.4)[11.0]
Spain				Norway			
Males	12.8	9.7 (9.5) [9.7]	16.8(16.2)[16.9]	Males	5.8	3.4 (2.9) [3.1]	9.8 (8.7) [9.0]
Females	15.6	11.9(11.7)[11.7]	20.3(19.5)[19.6]	Females	7.9	4.9 (4.5) [4.7]	12.6(11.4) [12]
Finland				Poland			
Males	3.4	1.9 (1.5) [1.7]	5.9 (5.2) [5.4]	Males	14.7	10.5(11.8)[10.2]	20.2(17.6) [20]
Females	5.4	3.2 (2.6) [3.2]	8.9 (8.2) [8.3]	Females	15.5	11.1(12.4)[10.5]	21.6(18.6) [21]
France				Portugal			
Males	8.3	6.6 (6.5) [6.6]	10.5(10.2)[10.3]	Males	10.2	6.0 (5.2) [5.5]	16.6(15.2)[15.5]
Females	10.1	8.0 (7.9) [7.9]	12.6(12.2)[12.4]	Females	9.7	5.7 (4.9) [5.3]	15.8(14.5)[15.1]
Greece				Sweden			
Males	17.0	11.8(11.6)[11.2]	23.6(22.3)[22.8]	Males	5.1	3.0 (2.6) [2.6]	8.1 (7.6) [7.9]
Females	14.9	10.6(10.4)[10.1]	20.8(19.5)[20.2]	Females	6.0	3.3 (2.7) [3.0]	10.3 (9.2) [9.6]
Hungary				Slovakia			
Males	16.6	12.2(12.0)[12.0]	22.5(21.2)[22.1]	Males	7.0	3.7 (2.8) [3.2]	13.1(11.2)[12.3]
Females	13.1	9.0 (8.6) [8.6]	18.8(17.5)[18.1]	Females	6.1	3.4 (2.9) [3.1]	10.1 (9.3) [9.8]
Ireland				United Kingdom			
Males	9.4	4.7 (3.4) [3.8]	17.0(15.3)[16.1]	Males	9.8	6.3 (5.9) [6.3]	14.6(13.8)[14.2]
Females	5.1	2.4 (1.1) [2.0]	10.6 (9.0) [9.1]	Females	9.9	5.6 (4.8) [5.1]	17.2(15.1)[16.4]

Table 5.4: Estimates of at-persistent-risk-of-poverty rate for 2009 by gender and age group between 25 and 44 year old. 95% empirical likelihood confidence intervals. The standard confidence intervals are given between brackets. Rescaled bootstrap confidence intervals are given between squared brackets.

Country	Rate %	Lower %	Upper %	Country	Rate %	Lower %	Upper %
Austria				Iceland			
Males	2.1	0.5 (-0.5) [0.1]	6.5 (4.8) [5.3]	Males	3.6	1.4 (0.6)[0.9]	7.5 (6.5) [7.3]
Females	4.7	2.5 (1.9) [2.0]	8.4 (7.5) [7.7]	Females	6.4	2.8 (1.8)[2.0]	12.7(11.1)[11.7]
Belgium				Italy			
Males	5.6	3.0 (2.4) [2.6]	9.9 (8.9) [9.4]	Males	9.5	7.1 (7.1)[6.9]	12.8(12.0)[12.4]
Females	8.6	5.5 (5.1) [5.1]	12.8(12.1)[12.1]	Females	12.5	9.7 (9.9)[9.4]	15.9(15.1)[15.6]
Bulgaria				Lithuania			
Males	6.6	3.7 (3.1) [3.3]	11.0(10.1)[10.4]	Males	8.0	4.7 (4.1)[4.3]	12.9(12.0)[12.9]
Females	8.9	5.3 (4.6) [4.9]	14.3(13.1)[13.8]	Females	9.8	5.7 (5)[5.5]	16.2(14.6)[15.0]
Cyprus				Luxembourg			
Males	2.4	1.0 (0.6) [0.7]	4.9 (4.3) [4.4]	Males	6.4	4.4 (4)[4.2]	9.5 (8.9) [9.1]
Females	4.4	2.4 (2.0) [2.1]	7.3 (6.8) [7.1]	Females	10.6	7.7 (7.3)[7.3]	14.6(13.8)[13.9]
Czech Republic				Latvia			
Males	1.5	0.7 (0.3) [0.4]	3.3 (2.8) [3.0]	Males	10.3	6.1 (5.0)[5.4]	17.4(15.6)[15.3]
Females	3.5	2.1 (1.8) [1.9]	5.8 (5.3) [5.4]	Females	9.4	6.0 (5.6)[5.9]	14.3(13.2)[13.5]
Denmark				Malta			
Males	3.5	1.1 (-0.1) [0.7]	8.9 (7.0) [7.8]	Males	2.9	1.0(-0.1)[0.6]	7.8 (5.9) [6.1]
Females	4.1	1.6 (0.8) [1.2]	8.5 (7.3) [7.6]	Females	3.1	1.5 (1.0)[1.2]	5.8 (5.2) [5.3]
Estonia				Netherlands			
Males	7.4	4.1 (2.9) [3.5]	14.7(12.0)[13.1]	Males	5.2	1.9 (0.7)[1.3]	11.7 (9.8)[10.2]
Females	7.3	5.0 (4.5) [4.9]	10.7(10.2)[10.2]	Females	3.6	1.0(-0.2)[0.4]	9.3 (7.4) [8.1]
Spain				Norway			
Males	7.7	5.8 (5.5) [5.5]	10.2 (9.9)[10.1]	Males	4.1	2.4 (2.0)[2.1]	6.9 (6.2) [6.3]
Females	9.0	7.0 (6.8) [7.0]	11.5(11.2)[11.4]	Females	3.4	2.0 (1.7)[1.8]	5.5 (5.0) [5.0]
Finland				Poland			
Males	2.9	1.5 (1.2) [1.3]	5.1 (4.6) [4.9]	Males	8.6	5.9 (6.3)[5.3]	12.5(10.8)[12.1]
Females	2.4	1.3 (0.9) [1.0]	4.2 (3.8) [3.9]	Females	9.4	6.9 (7.9)[6.7]	12.5(11.0)[12.6]
France				Portugal			
Males	4.5	3.3 (3.2) [3.2]	6.0 (5.8) [6.0]	Males	7.1	4.2 (3.7)[3.7]	11.3(10.6)[10.9]
Females	7.3	5.8 (5.6) [5.7]	9.2 (8.9) [9.0]	Females	7.0	3.9 (3.2)[3.4]	12.0(10.8)[11.3]
Greece				Sweden			
Males	11.3	7.5 (6.7) [7.0]	18.3(16.0)[17.0]	Males	3.8	2.1 (1.7)[2.0]	6.4 (6.0) [6.2]
Females	16.8	12.4(12.4)[12.2]	22.3(21.2)[21.8]	Females	4.2	2.4 (2.1)[2.1]	6.8 (6.4) [6.7]
Hungary				Slovakia			
Males	7.0	5.0 (4.7) [4.9]	9.7 (9.3) [9.5]	Males	3.0	1.5 (1.2)[1.3]	5.1 (4.7) [4.8]
Females	8.3	6.0 (5.8) [5.7]	11.2(10.8)[10.7]	Females	3.8	2.1 (1.7)[1.9]	6.5 (5.9) [6.0]
Ireland				United Kingdom			
Males	0.5	0.1 (-0.3) [0.0]	1.8 (1.3) [1.6]	Males	5.2	2.6 (1.8)[2.2]	9.9 (8.6) [8.9]
Females	5.3	2.0 (0.3) [1.3]	12.8(10.3)[10.7]	Females	7.3	4.5 (4.1)[4.3]	11.2(10.5)[10.8]

Table 5.5: Estimates of at-persistent-risk-of-poverty rate for 2009 by gender and age group > 44 year-old. 95% empirical likelihood confidence intervals. The standard confidence intervals are given between brackets. Rescaled bootstrap confidence intervals are given between squared brackets.

Country	Rate %	Lower %	Upper %	Country	Rate %	Lower %	Upper %
Austria				Iceland			
Males	7.3	5.0 (4.6) [4.6]	10.8(10.1)[10.6]	Males	3.1	1.0 (-0.2) [0.5]	8.4 (6.5) [7.0]
Females	9.5	7.4 (7.2) [7.3]	12.3(11.9)[12.3]	Females	3.2	1.0 (0.1) [0.6]	7.6 (6.2) [6.3]
Belgium				Italy			
Males	8.4	6.0 (5.8) [5.7]	11.4(11.0)[11.1]	Males	11.8	9.7 (9.7) [9.5]	14.3(13.9)[14.1]
Females	11.1	8.4 (8.2) [8.2]	14.4(13.9)[14.4]	Females	14.1	12.1(11.9)[12.1]	16.4(16.3)[16.3]
Bulgaria				Lithuania			
Males	8.8	6.5 (6.3) [6.5]	11.6(11.3)[11.6]	Males	9.1	6.6 (6.4) [6.6]	12.2(11.7)[11.8]
Females	13.0	10.2 (9.9) [9.9]	16.4(16.1)[16.0]	Females	14.7	12.1(12.0)[12.0]	17.8(17.4)[17.6]
Cyprus				Luxembourg			
Males	14.4	11.4(11.3)[11.3]	17.9(17.6)[17.7]	Males	5.3	3.5 (3.1) [3.1]	8.4 (7.6) [7.9]
Females	21.8	18.3(18.2)[18.2]	25.8(25.4)[25.5]	Females	6.5	4.5 (4.1) [4.2]	9.6 (8.9) [8.9]
Czech Republic				Latvia			
Males	2.4	1.5 (1.3) [1.4]	3.8 (3.5) [3.6]	Males	17.1	12.9(12.4)[12.7]	23.0(21.9)[22.1]
Females	3.8	2.9 (2.9) [2.8]	4.9 (4.7) [4.8]	Females	25.0	20.6(20.5)[20.3]	31.2(29.4)[30.1]
Denmark				Malta			
Males	8.8	5.7 (5.2) [5.2]	13.5(12.5)[13.1]	Males	6.7	4.3 (3.6) [3.9]	11.3 (9.8)[10.2]
Females	9.4	6.3 (6.0) [6.1]	13.7(12.9)[12.9]	Females	8.8	6.3 (6.1) [6.1]	11.7(11.4)[11.4]
Estonia				Netherlands			
Males	12.8	9.8 (9.6) [9.9]	16.5(16.0)[16.3]	Males	5.5	3.3 (2.9) [3.0]	8.8 (8.2) [8.5]
Females	19.3	15.9(15.8)[15.7]	23.3(22.7)[23.2]	Females	6.3	3.7 (3.3) [3.4]	10.2 (9.3) [9.6]
Spain				Norway			
Males	11.2	9.3 (9.2) [9.3]	13.4(13.2)[13.3]	Males	2.4	1.2 (0.8) [0.9]	4.6 (4.0) [4.3]
Females	13.4	11.2(11.1)[11.0]	15.9(15.7)[15.8]	Females	8.1	4.9 (4.4) [4.3]	12.9(11.8)[12.3]
Finland				Poland			
Males	8.9	6.1 (5.7) [5.8]	13.2(12.0)[12.6]	Males	8.5	6.3 (7.0) [6.1]	11.3(10.1)[11.2]
Females	13.2	9.7 (9.5) [9.2]	17.5(16.8)[17.2]	Females	7.7	5.7 (6.4) [5.4]	10.2 (9.0) [9.9]
France				Portugal			
Males	5.8	4.7 (4.7) [4.7]	7.1 (6.9) [7.0]	Males	10.2	7.6 (7.4) [7.3]	13.6(13.1)[13.5]
Females	6.5	5.6 (5.5) [5.5]	7.6 (7.5) [7.7]	Females	12.4	9.7 (9.6) [9.4]	15.7(15.3)[15.5]
Greece				Sweden			
Males	14.1	11.0(10.8)[10.7]	18.4(17.4)[17.9]	Males	4.6	3.0 (2.8) [2.8]	6.6 (6.3) [6.4]
Females	14.0	11.1(11.0)[11.0]	17.8(17.0)[17.5]	Females	8.6	6.2 (6.0) [6.1]	11.7(11.2)[11.3]
Hungary				Slovakia			
Males	4.8	3.6 (3.4) [3.3]	6.4 (6.2) [6.2]	Males	4.2	2.5 (2.2) [2.4]	6.7 (6.2) [6.1]
Females	5.0	3.8 (3.7) [3.8]	6.4 (6.3) [6.3]	Females	5.7	4.1 (3.9) [4.0]	7.8 (7.5) [7.8]
Ireland				United Kingdom			
Males	6.1	3.8 (3.3) [3.6]	9.3 (8.9) [9.0]	Males	8.6	6.7 (6.6) [6.5]	11.0(10.7)[10.7]
Females	9.0	6.0 (5.3) [5.8]	13.4(12.7)[12.7]	Females	9.0	7.0 (6.9) [7.0]	11.5(11.1)[11.2]

5.4 Concluding remarks

The results presented here are applications of the ultimate cluster approach where the sampling primary units are treated as units. The coverage of standard confidence intervals can be poor with skewed variables, or cannot preserve the range of the parameter space, for instance, the negative lower bounds in the case of domains of age and gender. The proposed approach is simpler than linearisation, naturally includes calibration. The proposed approach allows the calculation of likelihood ratio confidence intervals which is not the case for the calibration approach.

Bootstrap confidence intervals is an alternative approach in order to derive confidence intervals but it is more computationally intensive, especially with calibration weights and improvement of accuracy of bootstrap confidence intervals requires a considerable number of iterations. The empirical likelihood confidence intervals for the persistent-risk-of-poverty rate tend to be have an upward shift, compared with the bootstrap and standard confidence intervals, this can be explain by the fact the empirical likelihood preserve the range of the parameter space.

6

Algorithms for obtaining weights and intervals in the empirical likelihood approach

6.1 Introduction

The results presented in Chapter 3 correspond to an implementation of simulation analyses using artificial populations proposed by Wu and Rao (2006), and in Chapter 4 the simulation studies included the Wu and Rao (2006) population, sugar cane harvest (Chambers and Dunstan, 1986) and Swedish municipalities data (Särndal et al., 1992), and in Chapter 5 the analyses were performed using the *persistent-risk-of-poverty rate* indicator from EU-SILC survey user database, 2009. In this chapter, explanation of the key computational algorithms are provided for the proposed empirical likelihood point estimators, and the construction of empirical likelihood confidence intervals for the parameter of interest in the beforehand mentioned data sets. The computational tasks require the estimate of $\widehat{m}_i$ and $\ell(\widehat{m})$. Note that, the crucial algorithm is the computation of the vector $\boldsymbol{\eta}$, which is such that Equation (3.2.2) holds. The computational codes written in the statistical software R (R Core Team, 2012) are remitted to Chapter 6.

6.2 Algorithms

The computational routines to determine empirical likelihood point estimators and confidence intervals depend in mainly in the Algorithm 6.1, which finds $\boldsymbol{\eta}$ using a modified Newton-Raphson method (Wu, 2005), where $\boldsymbol{\eta}$ is the solution to (3.2.5) given $\boldsymbol{c}_i$ and $\boldsymbol{C}$. The modification needs finding the vector solution $\boldsymbol{\eta}$ in a range such that $\widehat{m}_i^{-1} > 0$ for all i (Chen et al., 2002). The modification ensure that the concave function (3.2.1) converges to the global maximum (see Section 3.2 for an extended description).

Algorithm 6.1: Estimation of $\boldsymbol{\eta}$ using a modified Newton-Raphson procedure.

Step 1. Let $\boldsymbol{\eta}_0 = \boldsymbol{0}$, $k = 0$ and $\epsilon = 10^{-12}$.

Step 2. Compute $\widehat{\boldsymbol{\Delta}}\left(\boldsymbol{\eta}_0\right)$ and $\boldsymbol{C} - \boldsymbol{f}(\boldsymbol{\eta}_0)$.

Step 3. Solve $\widehat{\boldsymbol{\Delta}}\left(\boldsymbol{\eta}_k\right)\left(\boldsymbol{\eta} - \boldsymbol{\eta}_0\right) = \boldsymbol{C} - \boldsymbol{f}(\boldsymbol{\eta}_k)$, and define the solution as $\boldsymbol{\gamma}_k$.

Step 4. If $\| \widehat{\boldsymbol{\Delta}}\left(\boldsymbol{\eta}_k\right)\left[\boldsymbol{C} - \boldsymbol{f}(\boldsymbol{\eta}_k)\right] \| < \epsilon$; stop the procedure, return $\boldsymbol{\eta}_k$; otherwise continue with next step.

Step 5. While $minimum\big\{\pi_i + (\boldsymbol{\eta}_k - \boldsymbol{\gamma}_k)^{\mathsf{T}}\boldsymbol{c}_i\big\} \leqslant 0$, replace $\boldsymbol{\gamma}_k$ by $\boldsymbol{\gamma}_k/2$, otherwise continue with next step.

Step 6. Set $\boldsymbol{\eta}_{k+1} = \boldsymbol{\eta}_k - \boldsymbol{\gamma}_k$ and go to step 2.

The value of $\boldsymbol{\eta}$ is replaced in the following equation: $\widehat{m}_i = \left(\pi_i + \boldsymbol{\eta}^{\mathsf{T}}\boldsymbol{c}_i\right)$, then the θ_0 can be estimated by $\widehat{\theta}$ which is the solution of the estimating equation (3.1.1). In step 2 in Algorithm 6.1, the vectors $\boldsymbol{c}_i$, $\boldsymbol{C}$ are modified accordingly to the sampling design considered (PPS sampling, πPS sampling or Rao-Hartley-Cochran sampling).

The algorithm for computing the lower (or upper) bound of an empirical likelihood confidence interval, under sampling with unequal probability is described in Algorithm

6.2. The seeking algorithm includes a previous stage to avoid searching for the lower (or upper) bound at extremely distant values from $\widehat{\theta}$, the empirical likelihood estimator of θ_0, where convergence may not be reached and reduces the computation time. In step 1 of Algorithm 6.2, the quantities $\widehat{m}_i$ are estimated accordingly to the approach considered, step 7 is associated with $\widehat{\theta}$ and therefore the estimating equation (3.3.2) has to be adapted in accordance with the parameter of interest θ_0 which can be a mean, total, quantile or the *persistent-risk-of-poverty* rate.

Algorithm 6.2: Lower (or upper) bound of the empirical likelihood confidence interval.

Step 1. Find the values of the weights $\widehat{m}_i$ and $\widehat{\theta}$ (the empirical likelihood estimator of θ_0).

Step 2. Set $b_1 = \widehat{\theta}$ and $b_2 = \min(y_i)$ (or $b_2 = \max(y_i)$ for the upper bound).

Step 3. Set $k = 0$ and $leap = |b_1 - b_2|/\sharp p$, where $\sharp p$ is the number of partitions of interval $|b_1 - b_2|$.

Step 4. Set $k = k + 1$ and $\tau_0 = \widehat{\theta} - leap \times k$ (or $\tau_0 = \widehat{\theta} + leap \times k$ for the upper bound), find the solution to Equation (3.3.2) for $\widehat{G}_\pi(\tau_0)$ and replace it in the Equation (3.3.1) in order to obtain the estimator $\widehat{r}(\tau_0)$.

Step 5. If $\widehat{r}(\tau_0) > \chi^2_1(\alpha)$, set $b_1 = \widehat{\theta}$ and $b_2 = \widehat{\theta} - leap \times (k+1)$ (or $b_2 = \widehat{\theta} + leap \times (k+1)$ for the upper bound); otherwise go to step 4.

Step 6. If $|b_1 - b_2| < 10^{-6}$, stop and set the lower bound $LB = (b_1 + b_2)/2$ (or $UB = (b_1 + b_2)/2$ for the upper bound). Otherwise, $\tau = (b_1 + b_2)/2$.

Step 7. Find the solution to Equation (3.3.2) for $\widehat{G}_\pi(\tau)$ and replace it in the Equation (3.3.1) in order to obtain the estimator $\widehat{r}(\tau)$.

Step 8. If $\widehat{r}(\tau) > \chi^2_1(\alpha)$, set $b_2 = \tau$. Otherwise $b_1 = \tau$. Go to step 6.

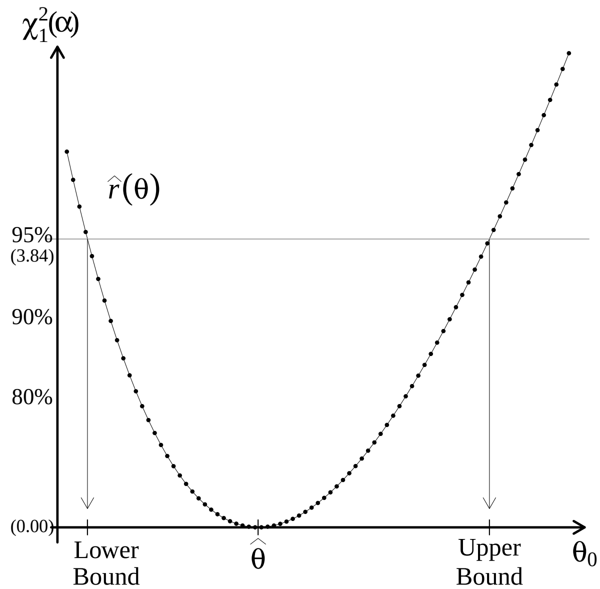

Figure 6.2: Empirical likelihood ratio confidence intervals for θ_0

Figure 6.2 shows the convexity of the empirical likelihood ratio function $\widehat{r}(\theta)$ which is evaluated against the cut-off value, 3.84, of the χ_1^2 distribution under the desired confidence level of $(1-\alpha)\%$ with $\alpha = 0.05$. The values of θ where $\widehat{r}(\theta) = 3.84$ define the upper bound and lower bound of the empirical likelihood confidence intervals.

7

Conclusions

Standard confidence intervals based on the central limit theorem and pseudo empirical likelihood confidence intervals require variance estimates which often involve linearisation or resampling. The coverage of standard confidence intervals can be poor with skewed variables or when the sampling distribution is not normal. The coverage and the tail errors can be also different from their intended levels. Even if the parameter of interest is not linear, the proposed method does not rely on normality of the point estimator, variance estimates, linearisation, resampling, and joint inclusion probabilities. Empirical likelihood confidence intervals can be easier to compute than standard confidence intervals based on variance estimates. The coverage of the proposed approach is usually better, as empirical likelihood confidence intervals are determined by the distribution of the data and the range of the parameter space is preserved.

There is an analogy between the proposed empirical likelihood approach and calibration (Huang and Fuller, 1978; Deville and Särndal, 1992), as the function (3.2.1) can be viewed as a calibration objective function, and the empirical likelihood estimator is asymptotically equivalent to a calibrated regression estimator (3.3.12). The objective functions used for calibration are disconnected from mainstream statistical theory. However, the proposed objective function (3.2.1) is related to the concept of likelihood. The advantage of the proposed empirical likelihood approach over standard calibration is the fact that (3.2.1) can be used to construct likelihood ratio confidence intervals. Empirical likelihood approaches are more general than calibration, and can be used for a wider class of parameters. The bootstrap is an alternative approach which can be used to derive non-parametric confidence intervals. The accuracy of the bootstrap confidence intervals has only been shown theoretically in a few particular cases (Rao and Wu, 1988). The proposed approach is simpler to implement and less computationally intensive than the bootstrap, especially with calibration weights. Our simulation study also shows that bootstrap confidence intervals may not have the right coverage and may be more unstable. From a practical point of view, the bootstrap is usually preferred because it does not rely on analytic derivation. The proposed approach possesses the same property, as it does not rely on any analytic derivation. Unlike the pseudo empirical likelihood approach, the proposed approach does not rely on variance estimates which could be difficult to estimate for complex parameters. This means that the proposed approach can be applied to a wider class of parameters. The proposed approach is also simpler to implement than the pseudo empirical likelihood. The simulation shows that the proposed approach may give better confidence intervals.

When the sample size is large, the Woodruff approach gives confidence intervals which are as good as the empirical likelihood confidence intervals in term of coverage and stability of the confidence intervals. However, the Woodruff approach relies on variance estimates and joint inclusion probabilities. Furthermore, this approach is only designed for quantiles. As mentioned before, the empirical likelihood approach can be used for a wider class of point estimators. The results presented here for the estimation of confidence intervals for the EU-SILC data are application of the ultimate cluster approach where the primary sampling units are treated as units. The coverage of standard confidence intervals can be poor with skewed variables, or cannot preserve the range of the parameter space, for instance, the negative lower bounds in the case of domains of age and gender. The proposed approach is simpler than linearisation, and naturally includes calibration. The proposed approach allows the calculation of likelihood ratio confidence intervals which is not the case for the calibration approach. Bootstrap confidence intervals is an alternative approach in order to derive confidence intervals but it is more computationally intensive, especially with calibration weights. An improvement of accuracy of bootstrap confidence intervals requires a considerable increasing in the number of iterations. The empirical likelihood confidence intervals for the persistent-risk-of-poverty rate tend to be have an upward shift, compared with the bootstrap and standard confidence intervals. This can be explained by the fact the empirical likelihood confidence intervals preserve the range of the parameter space.

Generalisation of the proposed empirical likelihood to multi-stage sampling with large fraction, estimation of confidence intervals for quantiles using auxiliary information, non-response, samples with relatively small size are areas of future research.

Appendix A

R functions for the empirical likelihood confidence intervals for survey data

A.1 Description of functions

In this appendix are described the main functions, written in the statistical software R (R Core Team, 2012), used in the simulation studies of this thesis. Only the functions for the implementation of the proposed empirical likelihood approach are considered. Functions for the implementation of compared approaches (standard methods based on normal distribution of data, pseudo empirical likelihood and bootstrap methods) are omitted. For these original papers are cited in the thesis. In this appendix, the functions related to the estimation of confidence intervals are limited only for the lower bound limit, the upper bound limit functions are similar, only the direction for the seeking should be inverted as it is indicated in Algorithm 6.2. The functions including a short description are ordered alphabetically in the following list.

Function	Description
`EST.EQ.MEAN`	Defines an estimating equation whose unique solution is the population mean (see Section 3.1).
`EST.EQ.QTILE`	Defines an estimating equation whose unique solution is the population quantile q, $0 < q < 1$ (see Section 4.1).
`ETA`	Computes the vector of $\boldsymbol{\eta}$ using the modified Newton-Raphson method (Wu, 2005; Polyak, 1987, see Section 3.2.2).
`LB.MEAN.CALIB`	Computes the lower bound of the empirical likelihood confidence interval for the point estimator of a population mean when sampling fraction is negligible ($nN^{-1} \leq 10\%$, Cochran, 1977) and auxiliary information is available (see Section 3.7). Functions `ETA` and `LOGLm` are called first.
`LB.MEAN.CALIB.LSF`	Computes the lower bound of the empirical likelihood confidence interval for the point estimator of a population mean when sampling fraction is non-negligible ($nN^{-1} > 10\%$, Cochran, 1977)and auxiliary information is available (see Section 3.8). Functions `ETA` and `LOGLm` are called first.
`LOGLm`	Computes the empirical log-likelihood $\ell(\widehat{m})$ (see Section 3.2) derived from the sum of the logarithm of $\widehat{m}_i$. The estimation of the mass units $\widehat{m}_i$ is based on the vector $\boldsymbol{\eta}$, the output of `ETA` function.

Function	Description
`LOWER.BOUND.MEAN`	Computes the lower bound of the empirical likelihood confidence interval for the point estimator of a population mean when sampling fraction is negligible (see Section 3.5). Functions `ETA` and `LOGLm` are called first.
`LOWER.BOUND.MEAN.LSF`	Computes the lower bound of the empirical likelihood confidence interval for the point estimator of a population mean when sampling fraction is non-negligible (see Section 3.8). Functions `ETA` and `LOGLm` are called first.
`LOWER.BOUND.QUANTILE`	Computes the lower bound of the empirical likelihood confidence interval for the point estimator of a population quantile Y_q when sampling fraction is negligible (see Section 3.7 and 4.2). Functions `ETA` and `LOGLm` are called first.
`LOWER.BOUND.QUANTILE.LSF`	Computes the lower bound of the empirical likelihood confidence interval for the point estimator of a population quantile Y_q when sampling fraction is non-negligible (see Section 3.7 and 4.2). Functions `ETA` and `LOGLm` are called first.
`WEIGHTS.mi`	Computes the vector of values $\widehat{m}_i$ (see (3.2.5)). The estimation of the mass units $\widehat{m}_i$ is based on the vector $\boldsymbol{\eta}$. Function `ETA` is called first.

A.2 R codes for empirical likelihood functions

EST.EQ.MEAN—Defines an estimating equation whose unique solution is the population mean (see Section 3.1).

```
function(N,mat.c,n,Pi.s,phi)
{
# EST.EQ.MEAN
# N = population size
# mat.c = matrix which each column represents a vector of design
#         variables, auxiliary variables or constraint variables and
#         the variable of interest
# Pi.s = vector of inclusion probabilities
# phi = point estimator of the population mean
# The sample size n is defined by other function that uses EST.EQ.MEAN
# for the searching of confidence intervals
 gi<-mat.c[,ncol(mat.c)]-(phi*N/n)*Pi.s
 return(gi)
}
```

EST.EQ.QTILE— Defines an estimating equation whose unique solution is the population quantile q, $0 < q < 1$ (see Section 4.1).

```
function(N,mat.c,theta,qtile)
{
# EST.EQ.QTILE
# N = population size
# mat.c = matrix which each column represents a vector of design
#         variables, auxiliary variables or constraint variables
#         and the variable of interest
 yi.q<-mat.c[,ncol(mat.c)]
 yi.knot<-yi.q[1]-(yi.q[2]-yi.q[1])
 yi.q<-c(yi.knot,yi.q)
 vec.rho<-NULL
 for (i.q in 2:length(yi.q))
 {delta.lag<-0
 if(yi.q[i.q-1]<=theta){delta.lag<-1}
 delta<-0
 if(yi.q[i.q]<=theta){delta<-1}
```

```
#______________________________________________________________________
#eta.rho is an interpolation function of the step distribution
#function used to find a unique solution
#______________________________________________________________________
 eta.rho<-
 delta+
 (theta-yi.q[i.q-1])/(yi.q[i.q]-yi.q[i.q-1])*(delta.lag)*(1-delta)
 vec.rho<-c(vec.rho,(eta.rho-qtile))}
 return(vec.rho)
}
```

`ETA`— Computes the vector of $\boldsymbol{\eta}$ using the modified Newton-Raphson method (Wu, 2005; Polyak, 1987), where $\boldsymbol{\eta} = Nn^{-1}\boldsymbol{\lambda}$ (see Section 3.2.2).

```
function(N,mat.c,Pi.s,Vec.C)
{
# ETA
# compute the vector eta
# N = population size
# mat.c = matrix which each column represents a vector of design
#         variables, auxiliary variables or constraint variables and
#         the variable of interest
# Vec.C = vector of known quantities
# Pi.s = vector of inclusion probabilities
 mat.c<-as.matrix(mat.c)
 n<-length(Pi.s)
 eta<-0*Vec.C
 dif<-1
 tol<-1e-12
 while(dif>tol){
 f<-0*Vec.C
 f1<-0*Vec.C
 delta<-f%*%t(f)
 for(i in 1:n){
 mi<-as.numeric( Pi.s[i]+t(eta)%*%(mat.c[i,]))
 f1<-f1+( mat.c[i,]/(n*mi) )
 delta<-delta-( mat.c[i,]%*%t(mat.c[i,])/mi^2)) }
 f<-f1/N
```

```
 eta0<-solve(delta,(f-Vec.C/N),tol=1e-14)
 dif<-t(f-Vec.C/N)%*%(f-Vec.C/N)
 rule<-1
 while(rule>0){
 rule<-0
 if( min(Pi.s+t(eta-eta0)%*%t(mat.c))<=0)
 rule<-rule+1
 if(rule>0)
 eta0<-eta0/2 }
 eta<-eta-eta0}
 return(eta)
}
```

`LB.MEAN.CALIB`— Computes the lower bound of the empirical likelihood confidence interval for the point estimator of a population mean when sampling fraction is negligible ($nN^{-1} \leq 10\%$, Cochran, 1977) and auxiliary information is available (see Section 3.7). Functions `ETA` and `LOGLm` are called first.

```
function(Q.chisq,N,mat.c,Pi.s,Vec.C,theta)
{
# LB.MEAN.CALIB
# Estimation of the empirical likelihood confidence
# interval lower limit for negligible sampling fraction
# and auxiliary information available
# Q.chisq = alpha level of the empirical likelihood confidence interval
# N = population size
# mat.c = matrix which each column represents a vector of design
#         variables, auxiliary variables or constraint variables and
#         the variable of interest
# Pi.s = vector of inclusion probabilities
# Vec.C = vector of known quantities
# theta = point estimator parameter of interest
 cut<-qchisq(Q.chisq,1)
 n<-length(Pi.s)
#______________________________________________________________
# Estimating equation specified for a population mean
#______________________________________________________________
 EST.EQ.MEAN<-function(N,mat.c,n,Pi.s,phi)
```

```
{gi<-mat.c[,ncol(mat.c)]-(phi*N/n)*Pi.s;return(gi)}
eta.vector<-
ETA(N,mat.c[,-ncol(mat.c)],Pi.s,Vec.C[-nrow(Vec.C),])
mi.restricted<-
c(WEIGHTS.mi(N,mat.c[,-ncol(mat.c)],Pi.s,eta.vector))
zi.dot<-mat.c[,ncol(mat.c)-1]/(Pi.s*mi.restricted)
mat.c[,ncol(mat.c)-1]<-zi.dot
lam<-
ETA(N,mat.c[,-ncol(mat.c)],1/mi.restricted,
        Vec.C[-nrow(Vec.C),] )
LogLik<-
LOGLm(N,as.matrix(mat.c[,-ncol(mat.c)]),1/mi.restricted,
       Vec.C[-nrow(Vec.C),],lam)
t1<-theta
t2<-min(mat.c[,ncol(mat.c)])
leap<-(t1-t2)/10
k1<-0
elrt1<-0
stp<-1
while(stp<cut)
{k1<-k1+1
tau1<-theta-leap*k1
Vec.C1<-matrix(c(Vec.C[-nrow(Vec.C),],0))
gi1<-EST.EQ.MEAN(N,mat.c,n,Pi.s,tau1)
mat.gi1<-cbind(mat.c[,-ncol(mat.c)],gi1)
vec.lam1<-ETA(N,mat.gi1,1/mi.restricted,Vec.C1)
LogLik.Y1<-LOGLm(N,mat.gi1,1/mi.restricted,Vec.C1,vec.lam1)
elrt1<-(2*(LogLik-LogLik.Y1 ))
stp<-elrt1}
tau1<-theta-leap*(k1+1)
t2<-tau1
crit<-t1-t2
elrt<-0
while(crit>1e-06)
{tau<-(t1+t2)/2
Vec.C2<-matrix(c(Vec.C[-nrow(Vec.C),],0))
gi2<-EST.EQ.MEAN(N,mat.c,n,Pi.s,tau)
```

```
 mat.gi2<-cbind(mat.c[,-ncol(mat.c)],gi2)
 vec.lam2<-ETA(N,mat.gi2,1/mi.restricted,Vec.C2)
 LogLik.Y2<-LOGLm(N,mat.gi2,1/mi.restricted,Vec.C2,vec.lam2)
 elrt2<-(2*(LogLik-LogLik.Y2 ))
 if (elrt2>cut) t2<-tau
 if (elrt2<=cut) t1<-tau
 crit<-t1-t2}
 LB<-(t2+t1)/2
 return(LB)
}
```

`LB.MEAN.CALIB.LSF`— Computes the lower bound of the empirical likelihood confidence interval for the point estimator of a population mean when sampling fraction is negligible ($nN^{-1} > 10\%$, Cochran, 1977) and auxiliary information is available (see Section 3.7 and 3.8). Functions `ETA` and `LOGLm` are called first.

```
function(Q.chisq,N,mat.c,Pi.s,Vec.C,theta)
{
# LBMEAN.CALIB.LSF
# Estimation of the empirical likelihood confidence
# interval lower limit for non-negligible sampling fraction
# and auxiliary information available
# Q.chisq = alpha level of the empirical likelihood confidence interval
# N = population size
# mat.c = matrix which each column represents a vector of design
#         variables, auxiliary variables or constraint variables and
#         the variable of interest
# Pi.s = vector of inclusion probabilities
# Vec.C = vector of known quantities
# theta = point estimator parameter of interest
 cut<-qchisq(Q.chisq,1)
 n<-length(Pi.s)
#_______________________________________________________________
# Estimating equation specified for a population mean
#_______________________________________________________________
 EST.EQ.MEAN<-function(N,mat.c,n,Pi.s,phi)
 {gi<-mat.c[,ncol(mat.c)]-(phi*N/n)*Pi.s;return(gi)}
 eta.vector<-
```

```
 ETA(N,mat.c[,-ncol(mat.c)],Pi.s,Vec.C[-nrow(Vec.C),])
 mi.restricted<-
 c(WEIGHTS.mi(N,mat.c[,-ncol(mat.c)],Pi.s,eta.vector))
 zi.dot<-mat.c[,ncol(mat.c)-1]/(Pi.s*mi.restricted)
 mat.c[,ncol(mat.c)-1]<-zi.dot
#_________________________________________________________________
# qi reduce the effect on the confidence intervals of units
# with large inclusion probabilities.
# ci.star.q and C.star.qi are the adjusted constraints by qi
#_________________________________________________________________
 qi<-sqrt(1-Pi.s)
 ci.star.qi<-qi*mat.c[,-ncol(mat.c)]
 C.star.qi<-matrix(t(mi.restricted)%*%ci.star.qi)
 eta<-ETA(N,ci.star.qi,1/mi.restricted,C.star.qi)
 LogLik<-LOGLm(N,ci.star.qi,1/mi.restricted,C.star.qi,eta)
 t1<-theta
 t2<-min(mat.c[,ncol(mat.c)])
 leap<-(t1-t2)/10
 k1<-0
 elrt1<-0
 stp<-1
 while(stp<cut)
 {k1<-k1+1
 tau1<-theta-leap*k1
 gi1<-EST.EQ.MEAN(N,mat.c,n,Pi.s,tau1)
 ci.star.qi.vector1<-cbind(ci.star.qi,gi1*qi)
 C.star.qi.vector1<-rbind(C.star.qi,sum((qi-1)*gi1*mi.restricted))
 eta.star.qi.vector1<-
 ETA(N,ci.star.qi.vector1,1/mi.restricted,C.star.qi.vector1)
 LogLik.Y1<-
 LOGLm(N,ci.star.qi.vector1,1/mi.restricted,
 C.star.qi.vector1,eta.star.qi.vector1)
 elrt1<-(2*(LogLik-LogLik.Y1))
 stp<-elrt1}
 tau1<-theta-leap*(k1+1)
 t2<-tau1
 crit<-t1-t2
```

```
 elrt<-0
 while(crit>1e-06)
 {tau<-(t1+t2)/2
 gi2<-EST.EQ.MEAN(N,mat.c,n,Pi.s,tau)
 ci.star.qi.vector2<-cbind(ci.star.qi,gi2*qi)
 C.star.qi.vector2<-rbind(C.star.qi,sum((qi-1)*gi2*mi.restricted))
 eta.star.qi.vector2<-
 ETA(N,ci.star.qi.vector2,1/mi.restricted,C.star.qi.vector2)
 LogLik.Y2<-LOGLm(N,ci.star.qi.vector2,1/mi.restricted,
 C.star.qi.vector2,eta.star.qi.vector2)
 elrt2<-(2*(LogLik-LogLik.Y2))
 if (elrt2>cut) t2<-tau
 if (elrt2<=cut) t1<-tau
 crit<-t1-t2}
 LB<-(t2+t1)/2
 return(LB)
}
```

`LOGLm`— Computes the empirical log-likelihood $\ell(\widehat{m})$ (see Section 3.2) derived from the sum of the logarithm of $\widehat{m}_i$. The estimation of the mass units $\widehat{m}_i$ is based on the vector $\boldsymbol{\eta}$, the output of `ETA` function.

```
function(N,mat.c,Pi.s,Vec.C,Vec.Eta)
{
# LOGLm
# Compute the empirical log-likelihood of the mass units using the
# output of ETA.
# N = population size
# mat.c = matrix which each column represents a vector of design
#         variables, auxiliary variables or constraint variables and
#         the variable of interest
# Vec.C = vector of known quantities
# Pi.s = vector of inclusion probabilities
# Vec.Eta = vector obtained from the ETA function
 mat.c<-as.matrix(mat.c)
 Vec.C<-t(cbind(Vec.C,deparse.level = 0))
 Pi.s<-cbind(Pi.s,deparse.level = 0)
 n<-length(Pi.s)
```

```
 L<-Vec.Eta
 dif<-1
 ac<-0
 for(i in 1:n){
        aa<-as.numeric(Pi.s[i]+t(L)%*%mat.c[i,])
        ab<-Pi.s[i]/(n*aa)
        ac<-ac+ab}
 af<-0
 for(i in 1:n){
       aa<-as.numeric(Pi.s[i]+t(L)%*%mat.c[i,])
        ad<-log( (Pi.s[i])/(n*aa*ac))
       af<-af+ad }
 return(af)
}
```

`LOWER.BOUND.MEAN`— Computes the lower bound of the empirical likelihood confidence interval for the point estimator of a population mean when sampling fraction is negligible (see Section 3.5). Functions `ETA` and `LOGLm` are called first.

```
function(Q.chisq,N,mat.c,Pi.s,Vec.C,theta)
{
# LOWER.BOUND.MEAN
# Estimation of the empirical likelihood confidence
# interval lower limit for negligible sampling fraction
# Q.chisq = alpha level of the empirical likelihood confidence interval
# N = population size
# mat.c = matrix which each column represents a vector of design
#         variables, auxiliary variables or constraint variables and
#         the variable of interest
# Pi.s = vector of inclusion probabilities
# Vec.C = vector of known quantities
# theta = point estimator parameter of interest
 cut<-qchisq(Q.chisq,1)
 mat.c<-cbind(mat.c,deparse.level = 0)
 n<-length(Pi.s)
#__________________________________________________________
# Estimating equation specified for a population mean
#__________________________________________________________
```

```
EST.EQ.MEAN<-function(N,mat.c,n,Pi.s,phi)
{gi<-mat.c[,ncol(mat.c)]-(phi*N/n)*Pi.s;return(gi)}
t1<-theta
t2<-min(mat.c[,ncol(mat.c)])
lam <- ETA(N,as.matrix(mat.c[,-ncol(mat.c)]),
       Pi.s,Vec.C[-nrow(Vec.C),])
LogLik <- LOGLm(N,as.matrix(mat.c[,-ncol(mat.c)]),
          Pi.s,Vec.C[-nrow(Vec.C),],lam)
leap<-(t1-t2)/10
k1<-0
elrt1<-0
stp<-1
while(stp<cut)
{k1<-k1+1
tau1<-theta+leap*k1
Vec.C1<-rbind(Vec.C[-nrow(Vec.C),],0,deparse.level = 0)
gi1<-EST.EQ.MEAN(N,mat.c,n,Pi.s,tau1)
mat.gi1<-cbind(mat.c[,-ncol(mat.c)],gi1)
vec.lam1<-ETA(N,mat.gi1,Pi.s,Vec.C1)
LogLik.Y1<-LOGLm(N,mat.gi1,Pi.s,Vec.C1,vec.lam1)
elrt1<-(2*(LogLik-LogLik.Y1 ))
stp<-elrt1}
tau1<-theta-leap*(k1+1)
t2<-tau1
crit<-t1-t2
elrt<-0
while(crit>1e-06)
{tau<-(t1+t2)/2
Vec.C2<-rbind(Vec.C[-nrow(Vec.C),],0,deparse.level = 0)
gi2<-EST.EQ.MEAN(N,mat.c,n,Pi.s,tau)
mat.gi2<-cbind(mat.c[,-ncol(mat.c)],gi2)
vec.lam2<-ETA(N,mat.gi2,Pi.s,Vec.C2)
LogLik.Y2<-LOGLm(N,mat.gi2,Pi.s,Vec.C2,vec.lam2)
elrt2<-(2*(LogLik-LogLik.Y2 ))
if (elrt2>cut) t2<-tau
if (elrt2<=cut) t1<-tau
crit<-t1-t2}
```

```
 LB<-(t2+t1)/2
 return(LB)
}
```

`LOWER.BOUND.MEAN.LSF`— Computes the lower bound of the empirical likelihood confidence interval for the point estimator of a population mean when sampling fraction is non-negligible (see Section 3.8). Functions `ETA` and `LOGLm` are called first.

```
function(Q.chisq,N,mat.c,Pi.s,Vec.C,theta)
{
# LOWER.BOUND.MEAN.LSF
# Estimation of the empirical likelihood confidence
# interval lower limit for non-negligible sampling fraction
# Q.chisq = alpha level of the empirical likelihood confidence interval
# N = population size
# mat.c = matrix which each column represents a vector of design
#         variables, auxiliary variables or constraint variables and
#         the variable of interest
# Pi.s = vector of inclusion probabilities
# Vec.C = vector of known quantities
# theta = point estimator parameter of interest
 cut<-qchisq(Q.chisq,1)
 n<-length(Pi.s)
#___________________________________________________________
# Estimating equation specified for a population mean
#___________________________________________________________
 EST.EQ.MEAN<-function(N,mat.c,n,Pi.s,phi)
 {gi<-mat.c[,ncol(mat.c)]-(phi*N/n)*Pi.s;return(gi)}
 t1<-theta
 t2<-min(mat.c[,ncol(mat.c)])
#___________________________________________________________
# qi reduce the effect on the confidence intervals of units
# with large inclusion probabilities.
# ci.star.q and C.star.qi are the adjusted constraints by qi
#___________________________________________________________
 qi<-sqrt(1-Pi.s)
 ci.star.qi<-qi*mat.c[,-ncol(mat.c)]
 C.star.qi<-colSums(matrix(ci.star.qi/Pi.s))
```

```
eta<-ETA(N,ci.star.qi,Pi.s,C.star.qi)
LogLik<-LOGLm(N,ci.star.qi,Pi.s,C.star.qi,eta)
leap<-(t1-t2)/1000
k1<-0
elrt1<-0
stp<-1
while(stp<cut)
{k1<-k1+1
tau1<-theta-leap*k1
gi1<-EST.EQ.MEAN(N,mat.c,n,Pi.s,tau1)
ci.star.qi.vector1<-cbind(ci.star.qi,gi1*qi)
C.star.qi.vector1<-rbind(C.star.qi,sum((qi-1)*gi1/Pi.s) )
eta.star.qi.vector1<-ETA(N,ci.star.qi.vector1,
                               Pi.s,C.star.qi.vector1)
LogLik.Y1<-LOGLm(N,ci.star.qi.vector1,Pi.s,
                 C.star.qi.vector1,eta.star.qi.vector1)
elrt1<-(2*(LogLik-LogLik.Y1))
stp<-elrt1}
tau1<-theta-leap*(k1+1)
t2<-tau1
crit<-t1-t2
elrt<-0
while(crit>1e-06)
{tau<-(t1+t2)/2
gi2<-EST.EQ.MEAN(N,mat.c,n,Pi.s,tau)
ci.star.qi.vector2<-cbind(ci.star.qi,gi2*qi)
C.star.qi.vector2<-rbind(C.star.qi,sum((qi-1)*gi2/Pi.s) )
eta.star.qi.vector2<-ETA(N,ci.star.qi.vector2,
                               Pi.s,C.star.qi.vector2)
LogLik.Y2<-LOGLm(N,ci.star.qi.vector2,Pi.s,
                 C.star.qi.vector2,eta.star.qi.vector2)
   elrt2<-(2*(LogLik-LogLik.Y2))
if (elrt2>cut) t2<-tau
if (elrt2<=cut) t1<-tau
crit<-t1-t2}
LB<-(t2+t1)/2
return(LB)
```

```
}
```

`LOWER.BOUND.QUANTILE`— Computes the lower bound of the empirical likelihood confidence interval for the point estimator of a population quantile Y_q when sampling fraction is negligible (see Section 3.7 and 4.2). Functions `ETA` and `LOGLm` are called first.

```
function(Q.chisq,N,mat.c,Pi.s,Vec.C,theta,qtile)
{
#LOWER.BOUND.QUANTILE
# Estimation of the empirical likelihood confidence
# interval lower limit for negligible sampling fraction
# Estimation of the empirical likelihood confidence
# interval lower limit for negligible sampling fraction
# Q.chisq = alpha level of the empirical likelihood confidence interval
# N = population size
# mat.c = matrix which each column represents a vector of design
#         variables, auxiliary variables or constraint variables and
#         the variable of interest
# Pi.s = vector of inclusion probabilities
# Vec.C = vector of known quantities
# theta = point estimator parameter of interest
# qtile = quantile 0 < q < 1
 cut<-qchisq(Q.chisq,1)
 mat.c<-cbind(mat.c,deparse.level = 0)
 n<-length(Pi.s)
#___________________________________________________________
# Estimating equation specified for a population quantile
#___________________________________________________________
 EST.EQ.QTILE<-function(N,mat.c,theta,qtile)
 {yi.q<-mat.c[,ncol(mat.c)]
 yi.knot<-yi.q[1]-(yi.q[2]-yi.q[1])
 yi.q<-c(yi.knot,yi.q)
 vec.rho<-NULL
 for (i.q in 2:length(yi.q))
 {delta.lag<-0
 if(yi.q[i.q-1]<=theta){delta.lag<-1}
 delta<-0
```

```
 if(yi.q[i.q]<=theta){delta<-1}
 eta.rho<- delta+
 (theta-yi.q[i.q-1])/(yi.q[i.q]-yi.q[i.q-1])*(delta.lag)*(1-delta)
 vec.rho<-c(vec.rho,(eta.rho-qtile))
 }
 return(vec.rho)
}
 t1<-theta
 t2<-min(mat.c[,ncol(mat.c)])
 lam<-ETA(N,as.matrix(mat.c[,-ncol(mat.c)]),
 Pi.s,Vec.C[-nrow(Vec.C),])
 LogLik<- LOGLm(N,as.matrix(mat.c[,-ncol(mat.c)]),
 Pi.s,Vec.C[-nrow(Vec.C),],lam)
 leap<-(t1-t2)/100
 k1<-0
 elrt1<-0
 stp<-1
 while(stp<cut)
 {k1<-k1+1
 tau1<-theta-leap*k1
 Vec.C1<-rbind(Vec.C[-nrow(Vec.C),],0,deparse.level = 0)
 gi1<-EST.EQ.QTILE(N,mat.c,tau1,qtile)
 mat.gi1<-cbind(mat.c[,-ncol(mat.c)],gi1)
 vec.lam1<-ETA(N,mat.gi1,Pi.s,Vec.C1)
 LogLik.Y1<-LOGLm(N,mat.gi1,Pi.s,Vec.C1,vec.lam1)
 elrt1<-(2*(LogLik-LogLik.Y1 ))
 stp<-elrt1}
 tau1<-theta-leap*(k1+1)
 t2<-tau1
 crit<-t1-t2
 elrt<-0
 while(crit>1e-06)
 {tau<-(t1+t2)/2
 Vec.C2<-rbind(Vec.C[-nrow(Vec.C),],0,deparse.level = 0)
 gi2<-EST.EQ.QTILE(N,mat.c,tau,qtile)
 mat.gi2<-cbind(mat.c[,-ncol(mat.c)],gi2)
 vec.lam2<-ETA(N,mat.gi2,Pi.s,Vec.C2)
```

```
 LogLik.Y2<-LOGLm(N,mat.gi2,Pi.s,Vec.C2,vec.lam2)
 elrt2<-(2*(LogLik-LogLik.Y2 ))
 if (elrt2>cut) t2<-tau
 if (elrt2<=cut) t1<-tau
 crit<-t1-t2}
 LB<-(t2+t1)/2
 return(LB)
}
```

LOWER.BOUND.QUANTILE.LSF— Computes the lower bound of the empirical likelihood confidence interval for the point estimator of a population quantile Y_q when sampling fraction is non-negligible (see Section 3.7 and 4.2). Functions ETA and LOGLm are called first.

```
function(Q.chisq,N,mat.c,Pi.s,Vec.C,theta,qtile)
{
# LOWER.BOUND.QUANTILE.LSF
# Estimation of the empirical likelihood confidence
# interval lower limit for non-negligible sampling fraction
# Q.chisq = alpha level of the empirical likelihood confidence interval
# N = population size
# mat.c = matrix which each column represents a vector of design
#         variables, auxiliary variables or constraint variables and
#         the variable of interest
# Pi.s = vector of inclusion probabilities
# Vec.C = vector of known quantities
# theta = point estimator par
 cut<-qchisq(Q.chisq,1)
 n<-length(Pi.s)
#_____________________________________________________________
# Estimating equation specified for a population quantile
#_____________________________________________________________
 EST.EQ.QTILE<-function(N,mat.c,theta,qtile)
 { yi.q<-mat.c[,ncol(mat.c)]
 yi.knot<-yi.q[1]-(yi.q[2]-yi.q[1])
 yi.q<-c(yi.knot,yi.q)
 vec.rho<-NULL
 for (i.q in 2:length(yi.q))
```

```
 {delta.lag<-0
 if(yi.q[i.q-1]<=theta){delta.lag<-1}
 delta<-0
 if(yi.q[i.q]<=theta){delta<-1}
 eta.rho<-delta+
 (theta-yi.q[i.q-1])/(yi.q[i.q]-yi.q[i.q-1])*(delta.lag)*(1-delta);
 vec.rho<-c(vec.rho,(eta.rho-qtile)) }
 return(vec.rho) }
 t1<-theta
 t2<-min(mat.c[,ncol(mat.c)])
#______________________________________________________________
# qi reduce the effect on the confidence intervals of units
# with large inclusion probabilities.
# ci.star.q and C.star.qi are the adjusted constraints by qi
#______________________________________________________________
 qi<-sqrt(1-Pi.s)
 ci.star.qi<-qi*mat.c[,-ncol(mat.c)]
 C.star.qi<-colSums(matrix(ci.star.qi/Pi.s))
 eta<-ETA(N,ci.star.qi,Pi.s,C.star.qi)
 LogLik<-LOGLm(N,ci.star.qi,Pi.s,C.star.qi,eta)
 leap<-(t1-t2)/200
 k1<-0
 elrt1<-0
 stp<-1
 while(stp<cut)
 {k1<-k1+1
 tau1<-theta-leap*k1
 gi1<-EST.EQ.QTILE(N,mat.c,tau1,qtile)
 ci.star.qi.vector1<-cbind(ci.star.qi,gi1*qi)
 C.star.qi.vector1<-rbind(C.star.qi,sum((qi-1)*gi1/Pi.s) )
 eta.star.qi.vector1<-ETA(N,ci.star.qi.vector1,
                              Pi.s,C.star.qi.vector1)
 LogLik.Y1<-LOGLm(N,ci.star.qi.vector1,Pi.s,
                 C.star.qi.vector1,eta.star.qi.vector1)
 elrt1<-(2*(LogLik-LogLik.Y1))
 stp<-elrt1}
 tau1<-theta-leap*(k1+1)
```

```
 t2<-tau1
 crit<-t1-t2
 elrt<-0
 while(crit>1e-06)
 {tau<-(t1+t2)/2
 gi2<-EST.EQ.QTILE(N,mat.c,tau,qtile)
 ci.star.qi.vector2<-cbind(ci.star.qi,gi2*qi)
 C.star.qi.vector2<-rbind(C.star.qi,sum((qi-1)*gi2/Pi.s) )
 eta.star.qi.vector2<-ETA(N,ci.star.qi.vector2,
                                Pi.s,C.star.qi.vector2)
 LogLik.Y2<-LOGLm(N,ci.star.qi.vector2,Pi.s,
                   C.star.qi.vector2,eta.star.qi.vector2)
 elrt2<-(2*(LogLik-LogLik.Y2))
 if (elrt2>cut) t2<-tau
 if (elrt2<=cut) t1<-tau
 crit<-t1-t2}
 LB<-(t2+t1)/2
 return(LB)
}
```

`WEIGHTS.mi`— Computes the vector of values $\widehat{m}_i$ (see (3.2.5)). The estimation of the mass units $\widehat{m}_i$ is based on the vector $\boldsymbol{\eta}$. Function `ETA` is called first.

```
function(N,mat.c,Pi.s,Vect.Eta)
{
# WEIGHTS.mi
# Compute the vector of mass units using the output of ETA.
# N = population size
# mat.c = matrix which each column represents a vector of design
#         variables, auxiliary variables or constraint variables and
#         the variable of interest
# Pi.s = vector of inclusion probabilities
# Vect.Eta = vector obtained from the ETA function
mat.c<-cbind(mat.c,deparse.level = 0)
Pi.s<-cbind(Pi.s)
n<-length(Pi.s)
mi<-0*Pi.s
for(i in 1:n)
```

```
  {
   mi.b<-as.numeric(Pi.s[i]+t(Vect.Eta)%*%(mat.c[i,]) )
   mi[i,]<-1/mi.b
  }
return(mi)
}
```

Appendix B

Proofs of Lemmas B.1, B.2, B.3 B.4 and Theorem B.5

The proof of the Theorems 3.3.10 and 3.7.1 and Corollaries 3.5.1, 3.6.1, 3.8.1 and 3.8.2 require four lemmas and a theorem. In this appendix the sketch proofs of the lemmas and theorem are shown. The lemmas were first proved in Berger and De La Riva Torres (2012).

Lemma B.1. *Let* $\boldsymbol{\eta}$ *the solution of* (3.2.2) *for given* $\boldsymbol{c}_i$ *and* $\boldsymbol{C}$. *Under the regularity conditions* (3.3.4)-(3.3.6), *we have* $\|\boldsymbol{\lambda}\|= O_p(n^{-\frac{1}{2}})$ *where* $\boldsymbol{\lambda} = Nn^{-1}\boldsymbol{\eta}$.

Proof. We have that $\boldsymbol{\lambda} =\|\boldsymbol{\lambda}\| \; \boldsymbol{L}$, where $\boldsymbol{L}$ is a vector which is such that $\|\boldsymbol{L}\|= O_p(1)$ and $\|\boldsymbol{L}\|^{-1}= O_p(1)$. The Equation (3.2.5) implies that

$$\widehat{m}_i = \pi_i^{-1}\left[1 - v_i(1+v_i)^{-1}\right], \tag{B.1.1}$$

where $v_i = n(N\pi_i)^{-1}\boldsymbol{c}_i^{\mathsf{T}}\boldsymbol{\lambda}$. Constraint (3.2.2) implies $\sum_{i=1}^{n}\widehat{m}_i\boldsymbol{c}_i = \boldsymbol{C}$. By multiplying the left side of the last equation by $N^{-1}\boldsymbol{L}^{\mathsf{T}}$ and using Equation (B.1.1), we obtain that for all $\boldsymbol{L}$

$$\boldsymbol{L}\;\|\boldsymbol{\lambda}\|\;\boldsymbol{L}^{\mathsf{T}}\widetilde{\boldsymbol{S}}\boldsymbol{L} = -N^{-1}\boldsymbol{L}^{\mathsf{T}}(\widehat{\boldsymbol{C}}_\pi - \boldsymbol{C}). \tag{B.1.2}$$

where $\widetilde{\boldsymbol{S}} = -nN^{-2}\sum_{i=1}^{n}\pi_i^{-2}(1+v_i)^{-1}\boldsymbol{c}_i\boldsymbol{c}_i^{\mathsf{T}}$. Note that $-\widehat{\boldsymbol{S}} \leq -(1+\max|v_i|)\widetilde{\boldsymbol{S}}$, this inequality and Equation (B.1.2) imply

$$\begin{aligned} -\boldsymbol{L}^{\mathsf{T}}\widehat{\boldsymbol{S}}\boldsymbol{L} &\leq (1+\max|v_i|)N^{-1}\,\|\boldsymbol{\lambda}\|^{-1}\,\boldsymbol{L}^{\mathsf{T}}|\widehat{\boldsymbol{C}}_\pi - \boldsymbol{C}| \\ &\leq (1+M\,\|\boldsymbol{\lambda}\|)N^{-1}\,\|\boldsymbol{\lambda}\|^{-1}\,|\boldsymbol{L}^{\mathsf{T}}(\widehat{\boldsymbol{C}}_\pi - \boldsymbol{C})|, \end{aligned} \tag{B.1.3}$$

as $\max|v_i| \leq \|\boldsymbol{\lambda}\|\ M$, where $M = \max\{nN^{-1}\pi_i^{-1}\ \|\boldsymbol{c}_i\|\} = o_p(n^{\frac{1}{2}})$. Hence Equation (B.1.3) implies that

$$\|\boldsymbol{\lambda}\|\ \{-\boldsymbol{L}^{\mathsf{T}}\widehat{\boldsymbol{S}}\boldsymbol{L} - MN^{-1}|\boldsymbol{L}^{\mathsf{T}}(\widehat{\boldsymbol{C}}_\pi - \boldsymbol{C})|\} \leq N^{-1}|\boldsymbol{L}^{\mathsf{T}}(\widehat{\boldsymbol{C}}_\pi - \boldsymbol{C})|. \tag{B.1.4}$$

Using conditions (3.3.4) and (3.3.5) we have that $|MN^{-1}\boldsymbol{L}^{\mathsf{T}}(\widehat{\boldsymbol{C}}_\pi - \boldsymbol{C})| = o_p(1)$. Finally, inequality (B.1.4) and (3.3.6) imply that $\|\boldsymbol{\lambda}\|\ [O_p(1) + o_p(1)] = O_p(n^{-\frac{1}{2}})$, because $-\boldsymbol{L}^{\mathsf{T}}\widehat{\boldsymbol{S}}\boldsymbol{L}$ is bounded away from zero, as $-\widehat{\boldsymbol{S}}$ is positive definite. □

Lemma B.2. *Let* $\widehat{m}_i = (\pi_i + \boldsymbol{\eta}^{\mathsf{T}}\boldsymbol{c}_i)^{-1}$. *Assuming the regularity conditions* (3.3.3)-(3.3.8) *hold for* $\mathbf{c}_i$ *and* $\boldsymbol{C}$*, we have that*

$$\boldsymbol{\eta} = nN^{-2}\widehat{\boldsymbol{S}}^{-1}(\boldsymbol{C} - \widehat{\boldsymbol{C}}_\pi) + nN^{-1}\widehat{\boldsymbol{e}}, \tag{B.2.1}$$

where $\widehat{\boldsymbol{e}}$ *is such that* $\|\widehat{\boldsymbol{e}}\| = O_p(n^{-1})$.

Proof. We have

$$\widehat{\boldsymbol{C}}_\pi - \boldsymbol{C} = \sum_{i=1}^{n}\boldsymbol{c}_i v_i \pi_i^{-1} - \sum_{i=1}^{n}\boldsymbol{c}_i\theta_i\pi_i^{-1}, \tag{B.2.2}$$

where v_i and $\boldsymbol{\lambda}$ are defined in Lemma B.1 and $\theta_i = \widehat{m}_i\pi_i - 1 + v_i$. By using the definition of $\widehat{\boldsymbol{S}}$ in (3.3.9), we have that

$$-nN^{-2}\widehat{\boldsymbol{S}}^{-1}\sum_{i=1}^{n}\boldsymbol{c}_i v_i\pi_i^{-1} = \boldsymbol{\eta}. \tag{B.2.3}$$

Equations (B.2.2) and (B.2.3) imply that

$$-nN^{-2}\widehat{\boldsymbol{S}}^{-1}(\widehat{\boldsymbol{C}}_\pi - \boldsymbol{C}) = \boldsymbol{\eta} - nN^{-1}\widehat{\boldsymbol{e}},$$

where $\widehat{\boldsymbol{e}} = -N^{-1}\widehat{\boldsymbol{S}}^{-1}\sum_{i=1}^{n}\boldsymbol{c}_i\theta_i\pi_i^{-1}$ which implies

$$\|\widehat{\boldsymbol{e}}\| \leq N^{-1} \ \|\widehat{\boldsymbol{S}}^{-1}\| \sum_{i=1}^{n} \|\boldsymbol{c}_i\| \ |\theta_i|\pi_i^{-1}. \tag{B.2.4}$$

We have that

$$|\theta_i| = |(1+v_i)^{-1} - 1 + v_i| \leq \gamma^{-1}v_i^2 = \frac{n^2}{\gamma N^2 \pi_i^2}(\boldsymbol{c}_i^{\mathsf{T}}\boldsymbol{\lambda})^2 \leq \frac{n^2}{\gamma N^2 \pi_i^2} \ \|\boldsymbol{c}_i\|^2 \|\boldsymbol{\lambda}\|^2, \tag{B.2.5}$$

using the fact that $|(1+v_i)^{-1} - 1 + v_i| \leq \gamma^{-1}v_i^2$ when $|v_i + 1| \geq \gamma > 0$. By combining (B.2.4) and (B.2.5), we have

$$\|\widehat{\boldsymbol{e}}\| \leq n^{-3}\gamma^{-1} \ \|\widehat{\boldsymbol{S}}^{-1}\| \|\boldsymbol{\lambda}\|^2 \ n^{-1}N^{-3} \sum_{i=1}^{n} \|\boldsymbol{c}_i\|^3 \ \pi_i^{-3} = O_p(n^{-1}), \tag{B.2.6}$$

using the regularity conditions (3.3.7), (3.3.8) and Lemma B.1. □

Lemma B.3. *Assuming that the regularity conditions* (3.3.3)-(3.3.8) *hold for* $\boldsymbol{c}_i$ *and* $\boldsymbol{C}$*, we have*

$$-2[\ell(\widehat{m}) + \ell(\pi)] = (\widehat{\boldsymbol{C}}_\pi - \boldsymbol{C})^{\mathsf{T}}\widehat{\boldsymbol{\Sigma}}^{-1}(\widehat{\boldsymbol{C}}_\pi - \boldsymbol{C}) + 2\boldsymbol{\eta}^{\mathsf{T}}\boldsymbol{C} + o_p(1), \tag{B.3.1}$$

where $\widehat{\boldsymbol{\Sigma}} = -N^2 n^{-1}\widehat{\boldsymbol{S}}$*.*

Proof. We have that

$$-2[\log(\widehat{m}_i) + \log(\pi_i)] = 2\log(1+v_i). \tag{B.3.2}$$

Using (3.3.3), (3.3.5) and Lemma B.1 we obtain that $\max(|v_i| \ : \ i \in s) = o_p(1)$. We have that $\log(1+v_i) = v_i - v_i^2/2 + \varphi_i$, where for some finite $\kappa > 0$, and $Pr\left\{|\varphi_i| \leq \kappa |v_i|^3, i \in s\right\} \to 1$ (see Owen, 2001, Ch. 11.2). Equation (B.3.2) implies that

$$-2[\log(\widehat{m}_i) - \log(\pi_i)] = 2\sum_{i=1}^{n} v_i - \sum_{i=1}^{n} v_i^2 + 2\sum_{i=1}^{n} \varphi_i. \tag{B.3.3}$$

Using Lemma B.2, we have that

$$\begin{aligned}\sum_{i=1}^{n} v_i = \sum_{i=1}^{n} \frac{\boldsymbol{c}_i^{\mathsf{T}}}{\pi_i}\boldsymbol{\eta} &= (\widehat{\boldsymbol{C}}_\pi - \boldsymbol{C})^{\mathsf{T}}\boldsymbol{\eta} + \boldsymbol{\eta}^{\mathsf{T}}\boldsymbol{C} = -nN^{-2}\widetilde{\boldsymbol{C}}_\pi^{\mathsf{T}}\widehat{\boldsymbol{S}}^{-1}\widetilde{\boldsymbol{C}}_\pi + \boldsymbol{\eta}^{\mathsf{T}}\boldsymbol{C} + nN^{-1}\widetilde{\boldsymbol{C}}_\pi^{\mathsf{T}}\widehat{\boldsymbol{e}} \\ &= -nN^{-2}\widetilde{\boldsymbol{C}}_\pi^{\mathsf{T}}\widehat{\boldsymbol{S}}^{-1}\widetilde{\boldsymbol{C}}_\pi + \boldsymbol{\eta}^{\mathsf{T}}\boldsymbol{C} + o_p(1),\end{aligned} \tag{B.3.4}$$

where $\widetilde{\boldsymbol{C}}_\pi = \widehat{\boldsymbol{C}}_\pi - \boldsymbol{C}$. Using Lemma B.2, Equation (3.3.9) and (B.2.1) we have that

$$\sum_{i=1}^{n} v_i^2 = \boldsymbol{\eta}^{\mathsf{T}} \sum_{i=1}^{n} \frac{1}{\pi_i^2} \boldsymbol{c}_i \boldsymbol{c}_i^{\mathsf{T}} \boldsymbol{\eta} = n^{-1} N^2 \widetilde{\boldsymbol{C}}_\pi^{\mathsf{T}} \widehat{\boldsymbol{S}}^{-1} \widetilde{\boldsymbol{C}}_\pi + 2n^{-1} N \widehat{\boldsymbol{e}}^{\mathsf{T}} \widetilde{\boldsymbol{C}}_\pi - n \widehat{\boldsymbol{e}}^{\mathsf{T}} \widehat{\boldsymbol{S}} \widehat{\boldsymbol{e}}. \quad \text{(B.3.5)}$$

By substituting (B.3.4) and (B.3.5) in (B.3.3), we have that

$$-2[\log(\widehat{m}_i) - \log(\pi_i)] = -n^{-1} N^2 \widetilde{\boldsymbol{C}}_\pi^{\mathsf{T}} \widehat{\boldsymbol{S}}^{-1} \widetilde{\boldsymbol{C}}_\pi + 2\boldsymbol{\eta}^{\mathsf{T}} \boldsymbol{C} + 2 \sum_{i=1}^{n} \varphi_i + o_p(1), \quad \text{(B.3.6)}$$

as $|\widehat{\boldsymbol{e}}^{\mathsf{T}} \widehat{\boldsymbol{S}}^{-1} \widehat{\boldsymbol{e}}| = o_p(n^{-1})$ and $|\widehat{\boldsymbol{e}}^{\mathsf{T}} \widetilde{\boldsymbol{C}}_\pi| = o_p(Nn^{-1})$ as $\|\widehat{\boldsymbol{e}}\| = O_p(n^{-1})$ (see Lemma B.2) and as Conditions (3.3.4) and (3.3.7) hold. We also have that

$$|\sum_{i=1}^{n} \varphi_i| \leq \kappa \sum_{i=1}^{n} |v_i|^3 \leq \kappa \, \|\boldsymbol{\eta}\|^3 \sum_{i=1}^{n} \frac{\|\boldsymbol{c}_i\|^3}{\pi_i^3} = O_p(n^{-\frac{1}{2}}), \quad \text{(B.3.7)}$$

using Condition (3.3.8) and $Nn^{-1} \, \|\boldsymbol{\eta}\| = O_p(n^{-\frac{1}{2}})$ (see Lemma B.1). Equations (B.3.6) and (B.3.6) imply (B.3.1). □

Lemma B.4. *Let* $\boldsymbol{c}_i^* = \left(Nn^{-1}\psi_i \boldsymbol{z}_i^{\mathsf{T}}, \boldsymbol{b}_i^{\mathsf{T}}\right)^{\mathsf{T}}$ *and* $\boldsymbol{C}^* = \left(Nn^{-1}\sum_{i=1}^{n} \psi_i \boldsymbol{z}_i^{\mathsf{T}} \pi_i^{-1}, \boldsymbol{b}_i^{\bullet\mathsf{T}}\right)^{\mathsf{T}}$, *where the* ψ_i *are bounded by* $\mathbf{b}_i$ *and* $\boldsymbol{b}_i$ *are vectors of constants such that the regularity conditions* (3.3.4)-(3.3.8) *hold for* $\boldsymbol{c}_i^*$ *and* $\boldsymbol{C}^*$. *Let* $\boldsymbol{\eta}^*$ *be such that* $\sum_{i=1}^{n} \widehat{m}_i^* \boldsymbol{c}_i^* = \boldsymbol{C}^*$, *where* $\widehat{m}_i^* = (\pi_i + \boldsymbol{\eta}^{*\mathsf{T}} \boldsymbol{c}_i^*)^{-1} = [\pi_i(1 + v_i^*)]^{-1}$, *and* $v_i^* = \pi_i^{-1} \boldsymbol{c}_i^{\mathsf{T}*} \boldsymbol{\eta}^*$. *We have that*

i) $\boldsymbol{\eta}^{\mathsf{T}*} \boldsymbol{C}^* = 0$ *when* $\psi_i = \psi$ *for all* i, *where* $\psi > 0$.

ii) $|\boldsymbol{\eta}^{*\mathsf{T}} \boldsymbol{C}^*| = O_p(n^{-\frac{1}{2}})$, *when there exists* $i \neq j$ *such that* $\psi_i \neq \psi_j$.

Proof. Let $\widetilde{\boldsymbol{S}}^* = -nN^{-2} \sum_{i=1}^{n} \pi_i^{-2} (1 + v_i^*)^{-1} \boldsymbol{c}_i^* \boldsymbol{c}_i^{*\mathsf{T}}$. As (B.1.2) is true for all $\boldsymbol{L}$, we can replace $\boldsymbol{L}$ by $\left(\mathbf{1}_H^{\mathsf{T}}, \mathbf{0}^{\mathsf{T}}\right)^{\mathsf{T}}$, where $\mathbf{1}_H$ is a $H \times 1$ vector of ones. As $\boldsymbol{\lambda}^* = Nn^{-1} \boldsymbol{\eta}^*$ and $\widehat{\boldsymbol{C}}_\pi^* - \boldsymbol{C}^* = \left(\mathbf{0}_H^{\mathsf{T}}, \sum_{i=1}^{n} \left(\boldsymbol{b}_i \pi_i^{-1} - \boldsymbol{b}_i^{\bullet}\right)^{\mathsf{T}}\right)^{\mathsf{T}}$, where $\mathbf{0}_H$ is a $H \times 1$ vector of zeros, (B.1.2) gives $\left(\mathbf{1}_H^{\mathsf{T}}, \mathbf{0}^{\mathsf{T}}\right)^{\mathsf{T}} \widetilde{\boldsymbol{S}}^* Nn^{-1} \boldsymbol{\eta}^* = -N^{-1} \left(\mathbf{1}_H^{\mathsf{T}}, \mathbf{0}^{\mathsf{T}}\right)^{\mathsf{T}} \left(\widehat{\boldsymbol{C}}_\pi^* - \boldsymbol{C}^*\right) = 0$ or equivalently

$$Nn^{-1} \mathbf{1}_H^{\mathsf{T}} \sum_{i=1}^{n} \widehat{m}_i^* \pi_i^{-1} \psi_i^2 \boldsymbol{z}_i \boldsymbol{z}_i^{\mathsf{T}} \boldsymbol{\eta}^{*(z)} + \mathbf{1}_H^{\mathsf{T}} \sum_{i=1}^{n} \widehat{m}_i^* \pi_i^{-1} \psi_i \boldsymbol{z}_i \boldsymbol{b}_i^{\mathsf{T}} \boldsymbol{\eta}^{*(b)} = 0, \quad \text{(B.4.1)}$$

where $\boldsymbol{\eta}^{*(z)}$ is a sub-vector of the first H components of $\boldsymbol{\eta}^*$. The coefficient $\boldsymbol{\eta}^{*(b)}$ is the last components of $\boldsymbol{\eta}^*$. Let $\eta_h^{*(z)}$ is the h-th component of $\boldsymbol{\eta}^*$. When $\mathbf{1}_H^\top \boldsymbol{z}_i = \pi_i$ Equation (B.4.1) reduces to

$$Nn^{-1}\sum_{h=1}^{H}\eta_h^{*(z)}\sum_{i\in s_h}\psi_i^2\widehat{m}_i^*\pi_i + \sum_{i=1}^{n}\widehat{m}_i^*\psi_i\boldsymbol{b}_i^\top\boldsymbol{\eta}^{*(b)} = 0. \tag{B.4.2}$$

We have that

$$\boldsymbol{\eta}^{*\top}\boldsymbol{C}^* = Nn^{-1}\sum_{h=1}^{H}\eta_h^{*(z)} + \sum_{i\in s_h}\psi_i + \sum_{i=1}^{n}\boldsymbol{b}_i^{\bullet\top}\boldsymbol{\eta}^{*(b)}. \tag{B.4.3}$$

When $\psi_i = \psi$ for all i, Equation (B.4.3) equals (B.4.2) divided by ψ because Constraint (3.2.2) implies $\sum_{i\in s_h}\psi\widehat{m}_i^*\pi_i = \sum_{i\in s_h}\psi$ and $\sum_{i=1}^n\widehat{m}_i^*\boldsymbol{b}_i = \sum_{i=1}^n\boldsymbol{b}_i^\bullet$. Thus $\boldsymbol{\eta}^{*\top}\boldsymbol{C}^* = 0$ and this proof the part $i)$ of the lemma.

By multiplying (B.4.2) by $\widehat{R} = [\sum_{i=1}^n\widehat{m}_i^*\boldsymbol{b}_i^\top\boldsymbol{\eta}^{*(b)}][\sum_{i=1}^n\widehat{m}_i^*\psi_i\boldsymbol{b}_i^\top\boldsymbol{\eta}^{*(b)}]^{-1}$ and the result subtracting to (B.4.3), we have

$$\begin{aligned}\boldsymbol{\eta}^{*\top}\boldsymbol{C}^* &= Nn^{-1}\sum_{h=1}^{H}\eta_h^{*(z)}\left(\sum_{i\in s_h}\psi_i - \widehat{R}\sum_{i\in s_h}\psi_i^2\widehat{m}_i^*\pi_i\right) + \left(\sum_{i=1}^{n}\boldsymbol{b}_i^{\bullet\top} - \sum_{i=1}^{n}\widehat{m}_i^*\boldsymbol{b}_i^\top\right)\\ &= Nn^{-1}\sum_{h=1}^{H}\eta_h^{*(z)}\left(\sum_{i\in s_h}\psi_i - \widehat{R}\sum_{i\in s_h}\psi_i^2\widehat{m}_i^*\pi_i\right),\end{aligned} \tag{B.4.4}$$

because $\sum_{i=1}^n\widehat{m}_i^*\boldsymbol{b}_i^{\bullet\top} = \sum_{i=1}^n\boldsymbol{b}_i^{\bullet\top}$ (see Constraint (3.2.2)). Then using the definition of $\widehat{R}$ and the fact that $\sum_{i\in s_h}\psi_i\widehat{m}_i^*\pi_i = \sum_{i\in s_h}\psi_i$ (see Constraint (3.2.2)), we have that (B.4.4) implies

$$\begin{aligned}|\boldsymbol{\eta}^{*\top}\boldsymbol{C}^*| &= \frac{N}{n}\left|\sum_{h=1}^{H}\eta_h^{*(z)}\left(\sum_{i\in s_h}\psi_i\widehat{m}_i^*\pi_i - \widehat{R}\sum_{i\in s_h}\psi_i^2\widehat{m}_i^*\pi_i\right)\right|\\ &= \frac{N}{n}\left|\sum_{i=1}^{n}\widehat{m}_i^*\boldsymbol{b}_i^\top\boldsymbol{\eta}^{*(b)}\sum_{h=1}^{H}\eta_h^{*(z)}\widehat{D}_h\right| \le \frac{N}{n}\,\|\sum_{i=1}^{n}\widehat{m}_i^*\boldsymbol{b}_i^\top\|\|\boldsymbol{\eta}^{*(b)}\|\|\boldsymbol{\eta}^{*(z)}\|\,\left(\sum_{h=1}^{H}\widehat{D}_h^2\right)^{\frac{1}{2}},\end{aligned} \tag{B.4.5}$$

using Cauchy's inequality, where

$$\widehat{D}_h = \frac{\sum_{i\in s_h}\psi_i\widehat{m}_i^*\pi_i}{\sum_{i=1}^n\widehat{m}_i^*\boldsymbol{b}_i^\top\boldsymbol{\eta}^{*(b)}} - \frac{\sum_{i\in s_h}\psi_i^2\widehat{m}_i^*\pi_i}{\sum_{i=1}^n\widehat{m}_i^*\boldsymbol{b}_i^\top\boldsymbol{\eta}^{*(b)}}.$$

As $\widehat{m}_i^* \doteq \pi_i^{-1}$, Condition (3.3.4) implies that $N^{-1} \parallel \sum_{i=1}^n \widehat{m}_i^* \boldsymbol{b}_i^\mathsf{T} \parallel = O_p(n^{-\frac{1}{2}})$ and $\widehat{D}_h = O_p(1)$. Furthermore, $\parallel \boldsymbol{\eta}^{*(b)} \parallel \parallel \boldsymbol{\eta}^{*(z)} \parallel \leq \parallel \boldsymbol{\eta}^* \parallel^2$ and Lemma B.1 implies that $\|\boldsymbol{\eta}^{*(b)}\|\|\boldsymbol{\eta}^{*(z)}\| \leq O_p(nN^{-2})$. Thus, Equation (B.4.5) implies that $\|\boldsymbol{\eta}^{*\mathsf{T}}\boldsymbol{C}^*\| = O_p(n^{-\frac{1}{2}})$, as the number of strata is bounded. This shows part $ii)$ of the lemma. □

Theorem B.5. *Let* $\boldsymbol{c}_i^* \left(Nn^{-1}\psi_i \boldsymbol{z}_i, \psi_i^\bullet g_i(\theta_0)\right)^\mathsf{T}$ *and* $\boldsymbol{C}^* = \left(Nn^{-1}\sum_{i=1}^n \psi_i \breve{\boldsymbol{z}}_i^\mathsf{T}, \sum_{i=1}^n (\psi_i^\bullet - 1)\breve{g}_i(\theta_0)\right)^\mathsf{T}$, *where* ψ_i *and* $\psi_i^\bullet$ *are bounded. The values* ψ_i *and* $\psi_i^\bullet$ *are such that the regularity conditions* (3.3.4)-(3.3.8) *hold for* $\boldsymbol{c}_i = \boldsymbol{c}_i^*$ *and* $\boldsymbol{C} = \boldsymbol{C}^*$. *Let* $\widehat{m}_i^*(\theta_0) = \left(\pi_i + \boldsymbol{\eta}^{*\mathsf{T}}\boldsymbol{c}_i^*\right)^{-1}$ *and* $\ell(\widehat{m}^*, \theta_0) = \sum_{i=1}^n \log(\widehat{m}_i^*(\theta_0))$, *we have that*

$$-2[\ell(\widehat{m}^*, \theta_0) + \ell(\pi)] = \widehat{G}_\pi(\theta_0)^2 \left(\widehat{\sigma}_{gg} - \widehat{\boldsymbol{\Sigma}}_{zg}^\mathsf{T}\widehat{\boldsymbol{\Sigma}}_{zz}^{-1}\widehat{\boldsymbol{\Sigma}}_{zg}\right)^{-1} + o_p(1), \tag{B.5.1}$$

where $\widehat{\sigma}_{gg} = \sum_{i=1}^n \psi_i^{\bullet 2}\breve{g}_i(\theta_0)^2$, $\widehat{\boldsymbol{\Sigma}}_{zz} = N^{-2}n^{-2}\sum_{i=1}^n \psi_i^2 \breve{\boldsymbol{z}}_i\breve{\boldsymbol{z}}_i^\mathsf{T}$, $\widehat{\boldsymbol{\Sigma}}_{zg} = Nn^{-1}\sum_{i=1}^n \psi_i^\bullet\psi_i\breve{\boldsymbol{z}}_i\breve{g}_i(\theta_0)$.

Proof. Using Lemma B.4 with $\boldsymbol{b}_i = \psi_i^\bullet g_i(\theta_0)$ and $\boldsymbol{b}_i^\bullet = (\psi_i^\bullet - 1)\breve{g}_i(\theta_0)$, we have that $|\boldsymbol{\eta}^{*\mathsf{T}}\boldsymbol{C}^*| = O_p(n^{-\frac{1}{2}})$. Thus, Lemma B.3 implies

$$-2[\ell(\widehat{m}^*, \theta_0) + \ell(\pi)] = (\widehat{\boldsymbol{C}}_\pi^* - \boldsymbol{C}^*)^\mathsf{T}\widehat{\boldsymbol{\Sigma}}^{*-1}(\widehat{\boldsymbol{C}}_\pi^* - \boldsymbol{C}^*) + o_p(1), \tag{B.5.2}$$

where $\widehat{\boldsymbol{\Sigma}}^* = -N^2n^{-1}\widehat{\boldsymbol{S}}$. The matrix $\widehat{\boldsymbol{\Sigma}}$ can be re-written as

$$\widehat{\boldsymbol{\Sigma}}^* = \sum_{i=1}^n \frac{1}{\pi_i}\boldsymbol{c}_i^*\boldsymbol{c}_i^{*\mathsf{T}} = \begin{pmatrix} \widehat{\boldsymbol{\Sigma}}_{zz} & \widehat{\boldsymbol{\Sigma}}_{zg} \\ \widehat{\boldsymbol{\Sigma}}_{zg}^\mathsf{T} & \widehat{\sigma}_{gg} \end{pmatrix}. \tag{B.5.3}$$

We have that $\widehat{\boldsymbol{C}}_\pi^* - \boldsymbol{C}^* = (\boldsymbol{0}_H^\mathsf{T}, \widehat{G}_\pi(\theta_0))^\mathsf{T}$. Hence (B.5.3) implies that

$$\begin{aligned} -2[\ell(\widehat{m}^*, \theta_0) + \ell(\pi)] &= (\boldsymbol{0}_H^\mathsf{T}, \widehat{G}_\pi(\theta_0))\widehat{\boldsymbol{\Sigma}}^{*-1}(\boldsymbol{0}_H^\mathsf{T}, \widehat{G}_\pi(\theta_0))^\mathsf{T} + o_p(1) \\ &= \widehat{G}_\pi(\theta_0)^2 \left(\widehat{\sigma}_{gg} - \widehat{\boldsymbol{\Sigma}}_{zg}^\mathsf{T}\widehat{\boldsymbol{\Sigma}}_{zz}^{-1}\widehat{\boldsymbol{\Sigma}}_{zg}\right)^{-1} + o_p(1), \end{aligned} \tag{B.5.4}$$

using Schur complement of $\widehat{\boldsymbol{\Sigma}}_{zz}$. □

References

Aires, N. (2000). Comparisons between conditional poisson sampling and Pareto πps sampling designs. *Journal of Statistical Planning and Inference*, 82:1–15.

Antal, E. and Tillé, Y. (2010). Simple random sampling with over-replacement. *Journal Statistical Planning Inference*, 141(1):597–601.

Antal, E. and Tillé, Y. (2011). A direct bootstrap method for complex sampling designs from a finite population. *Journal of the American Statistical Association*, 106(494):534–543.

Bahadur, R. R. (1966). A note on quantiles in large samples. *The Annals of Mathematical Statistics*, 37:388393.

Bansal, M. and Singh, R. (1986). On the generalization of Rao, Hartley and Cochran's scheme. *Metrika*, 33(1):307–314.

Berger, Y. (2005a). A variance estimator for systematic sampling from a deliberated ordered population. *Communications in Statistics-Theory and Methods*, 34:1533–1541.

Berger, Y. G. (1998). Rate of convergence to normal distribution for the Horvitz-Thompson estimator. *Journal of Statistical Planning and Inference*, 67:209–226.

Berger, Y. G. (2005b). Variance estimation with Chao's sampling scheme. *Journal of Statistical Planning and Inference*, 127(12):253 – 277.

Berger, Y. G. (2011). Asymptotic consistency under large entropy sampling designs with unequal probabilities. *Pakistan Journal of Statistics, Festschrift to honour Ken Brewer's 80th birthday*, 27:407–426.

Berger, Y. G. and De La Riva Torres, O. (2012). A unified theory of empirical likelihood ratio confidence intervals for survey data with unequal probabilities and non negligible sampling fractions. *Southampton Statistical Sciences Research Institute http://eprints.soton.ac.uk/337688/*. 24pp.

Berger, Y. G. and Skinner, C. J. (2003). Variance estimation of a low-income proportion. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52:457–468.

Berger, Y. G., Tirari, M. E. H., and Tillé, Y. (2003). Towards optimal regression estimation in sample surveys. *Australian and New Zealand Journal of Statistics*, 45:319–329.

Binder, D. A. and Kovacevìc, M. S. (1995). Estimating some measure of income inequality from survey data: an application of the estimating equation approach. *Survey Methodology*, 21:137–145.

Brewer, K. R. W. (2002). *Combined Survey Sampling Inference*. Arnold, London.

Brewer, K. R. W. and Donadio, M. E. (2003). The high entropy variance of the Horvitz-Thompson estimator. *Survey Methodology*, 29:189–196.

Chambers, R. L., R. L. and Dunstan, R. (1986). Estimating distribution functions from survey data. *Biometrika*, 73(3):597–604.

Chao, M. T. (1982). A general purpose unequal probability sampling plan. *Biometrika*, 69(3):653–656.

Chen, J., Chen, S. Y., and Rao, J. N. K. (2003). Empirical likelihood confidence intervals for the mean of a population containing many zero values. *The Canadian Journal of Statistics*, 31(1):53–68.

Chen, J. and Qin, J. (1993). Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. *Biometrika*, 80(1):107–116.

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.

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(1):230–237.

Chen, S. X. and Hall, P. (1993). Smoothed empirical likelihood confidence intervals for quantiles. *The Annals of Statistics*, 21(3):1166–1181.

Chen, S. X. and Qin, J. (2003). Empirical likelihood-based confidence intervals for data with possible zero observations. *Statistics & Probability Letters*, 65:29–37.

Cochran, W. (1977). *Sampling techniques, 3rd Edition*. John Wiley and Sons, New York.

Deming, W. E. and Stephan, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Ann. Math. Statist.*, 11(4):427–444.

Deville, J. C. (1999). Variance estimation for complex statistics and estimators: linearization and residual techniques. *Survey Methodology*, 25:193–203.

Deville, J. C. and Särndal, C. E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87(418):376–382.

Deville, J.-C. and Tillé, Y. (2005). Variance approximation under balanced sampling. *Journal of Statistical Planning and Inference*, 128:569–591.

Dorfman, A. H. (2009). *Sample Surveys: Inference and Analysis*, volume 29B of *Handbook of statistics*, chapter Inference on Distribution Functions and Quantiles, pages 371–395. North-Holland, The Netherlands.

Durbin, J. (1953). Some results in sampling theory when the units are selected with unequal probabilities. *Journal of the Royal Statistical Society Series B*, 15:262–269.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statis-*

tics, 7:1–26.

Eurostat (2012). European union statistics on income and living conditions (EU-SILC). http://epp.eurostat.ec.europa.eu/portal/page/portal/microdata/eu_silc. [Online; accessed 8 July 2013].

Francisco, C. A. and Fuller, W. A. (1991). Quantile estimation with a complex survey design. *The Annals of Statistics*, pages 454–469.

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

Godambe, V. (1966). A new approach to sampling from finite population i, ii. *Journal of the Royal Statistical Society Series B*, 28:310–328.

Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *The Annals of Mathematical Statistics*, 35(35):1491–1523.

Hájek, J. (1971). *Foundations of Statistical Inference*, chapter Comment on An essay on the logical foundations of survey sampling by D. Basu. Toronto: Holt and Rinehart and Winston.

Hájek, J. (1981). *Sampling from a finite population*. Marcel Dekker, New York.

Hall, P. (1992). *The Bootstrap and Edgeworth expansion*. Springer-Verlag, New York.

Hall, P. and La Scala, B. (1990). Methodology and algorithms of empirical likelihood. *International Statistical Review*, 58(2):109–127.

Hampel, F. R., Rosseeuw, P. J., Ronchetti, E. M., and Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. John Wiley and Sons, New York.

Hansen, M., Hurwitz, W., and Madow, W. (1953). *Sample Survey Methods and Theory*, volume I. John Wiley and Sons, New York.

Hansen, M. H. and Hurwitz, W. N. (1943). On the theory of sampling from finite populations. *The Annals of Mathematical Statistics*, 17:333–362.

Hartley, H. O. and Rao, J. N. K. (1968). A new estimation theory for sample surveys. *Biometrika*, 55(3):547–557.

Hartley, H. O. and Rao, J. N. K. (1969). *A new estimation theory for sample surveys, II*. Wiley-Interscience, New York.

Haziza, D., Mecatti, F., and Rao, J. N. K. (2004). Comparison of variance estimators under Rao-Sampford method: a simulation study. *Proceedings of the Survey Methods Section, American Statistical Association.*

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685.

Huang, E. T. and Fuller, W. A. (1978). Nonnegative regression estimation for survey data. *Proceedings Social Statistics Section American Statistical Association*, pages 300–303.

Hudson, D. J. (1971). Interval estimation from the likelihood function. *Journal of the Royal Statistical Society Series B*, 33:256–262.

Isaki, C. T. and Fuller, W. A. (1982). Survey design under the regression superpopulation model. *Journal of the American Statistical Association*, 77:89–96.

Jing, B.-Y., Yuan, J., and Zhou, W. (2009). Jackknife empirical likelihood. *Journal of the American Statistical Association*, 104:1224–1232.

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

Kott, P. (2009). *Sample Surveys: Inference and Analysis*, volume 29B of *Handbook of statistics*, chapter Calibration Weighting: Combining Probability Samples and Linear Prediction Models, pages 55–82. North-Holland, The Netherlands.

Kovar, J. G., Rao, J. N. K., and Wu, C. F. J. (1988). Bootstrap and other methods to measure errors in survey estimates. *Canadian Journal of Statistics*, 16(S1):25–45.

Krewski, D. and Rao, J. N. K. (1981). Inference from stratified sample: properties of linearization jacknife, and balanced repeated replication methods. *The Annals of Statistics*, 9:1010–1019.

Lesage, E. (2011). The use of estimating equations to perform a calibration on complex parameters. *Survey Methodology*, 37:103–108.

Montanari, G. (1987). Post sampling efficient qr-predictions in large sample survey. *International Statistical Review*, 55:191–202.

Ohlsson, E. (1986). Normality of the Rao, Hartley, Cochran estimator: An application of the martingale clt. *Scandinavian Journal of Statistics*, 12:17–28.

Osier, G. (2009). Variance estimation for complex indicators of poverty and inequality using linearization techniques. In *Survey Research Methods*, volume 3, pages 167–195.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75(2):237–249.

Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics*, 18(1):90–120.

Owen, A. B. (1991). Empirical likelihood for linear models. *The Annals of Statistics*, 19:17251747.

Owen, A. B. (2001). *Empirical likelihood*. Chapman & Hall, New York.

Polyak, B. T. (1987). *Introduction to Optimization*. Optimization Software, Inc., Publications Division, New York.

Prášková, Z. and Sen, P. K. (2009). *Vol. 29B of Handbook of Statistics*, chapter Asymptotic in Finite Population Sampling, Sample Surveys: Inference and Analysis, pages 489–522. North-Holland, The Netherlands.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics*, pages 300–325.

R Core Team (2012). *R: A Language and Environment for Statistical Computing*. R

Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Rao, J. N. K. (1965). On two simples schemes of unequal probability sampling without replacement. *Journal of the Indian Statistical Association*, 3:173–180.

Rao, J. N. K. (1966). Alternative estimators in pps sampling for multiple characteristics. *Sankhyā*, A28:47–60.

Rao, J. N. K. (1994). Estimating total and distribution function using auxiliary information at the estimation stage. *Journal of Official Statistics*, 10(2):153–165.

Rao, J. N. K., Hartley, H. O., and Cochran, W. G. (1962). On a simple procedure of unequal probability sampling without replacement. *Journal of the Royal Statistical Society Series B*, 24:482–491.

Rao, J. N. K. and Wu, C. (2009). *Sample Surveys: Inference and Analysis*, volume 29B of *Handbook of statistics*, chapter Empirical Likelihood Methods, pages 189–208. North-Holland, The Netherlands.

Rao, J. N. K. and Wu, C. F. J. (1988). Resampling inference with complex survey data. *Journal of the American Statistical Association*, 83(401):231–241.

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.

Salehi, M., Mohammadi, M., Rao, J. N. K., and Berger, Y. (2010). Empirical likelihood confidence intervals for adaptive cluster sampling. *Environmental and Ecological Statistics*, 17:111–123.

Sampford, M. R. (1967). On sampling without replacement with unequal probabilities of selection. *Biometrika*, 54(3/4):499–513.

Särndal, C. E. (1996). Efficient estimators with simple variance in unequal probability sampling. *Journal of the American Statistical Association*, 91:1289–1300.

Särndal, C. E. (2007). The calibration approach in survey theory and practice. *Survey Methodology*, 33(2):99–119.

Särndal, C.-E., Swensson, B., and Wretman, J. (1992). *Model assisted survey sampling*. Springer-Verlag, New York.

Scott, A. and Wu, C. F. (1981). On the asymptotic distribution of the ratio and regression estimator. *Journal of the American Statistical Association*, pages 98–102.

Sen, P. K. (1953). On the estimate of the variance in sampling with varying probabilities. *Journal of the Indian Society of Agricultural Statistics*, 5:119–127.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

Tillé, Y. (2006). *Sampling Algorithms*. Springer Series in Statistics. Springer, New York.

Vísek, J. (1979). Asymptotic distribution of simple estimate for rejective, sampford and successive sampling. *Contribution to Statistics, Jaroslav Hájek Memorial Volume. Academia of Prague, Czech Republic*, pages 71–78.

Wilks, S. S. (1938). Shortest average confidence intervals from large samples. *The Annals of Mathematical Statistics*, 9:166–175.

Woodruff, R. S. (1952). Confidence intervals for medians and other position measures. *Journal of the American Statistical Association*, 47:635–646.

Wu, C. (2004). R/S-plus implementation of pseudo empirical likelihood methods under unequal probability sampling. Technical Report 07, Department of Statistics and Actuarial Science, University of Waterloo.

Wu, C. (2005). Algorithms and R codes for the pseudo empirical likelihood method in survey sampling. *Survey Methodology*, 31(2):239–243.

Wu, C. and Rao, J. N. K. (2006). Pseudo-empirical likelihood ratio confidence intervals for complex surveys. *The Canadian Journal of Statistics*, 34(3):359–375.

Yates, F. and Grundy, P. M. (1953). Selection without replacement from within strata with probability proportional to size. *Journal of the Royal Statistical Society Series B*, 1:253–261.

Zhong, B. and Rao, J. N. K. (1996). Empirical likelihood inference under stratified random sampling using auxiliary information. *ASA Proceedings of the Section on Survey Research Methods*, 87:798–803.

Zhong, B. and Rao, J. N. K. (2000). Empirical likelihood inference under stratified random sampling using auxiliary population. *Biometrika*, 87(4):929–938.